

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:14:15 ; Search time 30 Seconds  
(without alignments)  
2503.232 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 5273  
Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAFPVSVFVHTAD 1006

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5253	99.6	1006	US-09-023-905A-12	Sequence 12, Appl
2	3874	73.5	982	US-09-023-905A-7	Sequence 7, Appl
3	3155	59.8	903	US-09-023-905A-10	Sequence 10, Appl
4	3013.5	57.1	1129	US-09-023-905A-2	Sequence 2, Appl
5	2954.5	56.0	1151	US-09-023-905A-4	Sequence 4, Appl
6	1387	26.3	307	US-09-023-905A-36	Sequence 36, Appl
7	484.5	9.2	134	US-09-023-905A-30	Sequence 30, Appl
8	453	8.6	166	US-09-270-767-33258	Sequence 33258, A
9	419	7.9	605	US-09-828-303-19	Sequence 19, Appl
10	347.5	6.6	225	US-09-270-767-42748	Sequence 42748, A
11	278	5.3	802	US-09-949-016-6235	Sequence 6235, Ap
12	262	5.0	750	US-09-949-016-11166	Sequence 11166, A
13	261.5	5.0	913	US-09-248-796A-15734	Sequence 15734, A
14	223	4.2	377	US-09-270-767-42077	Sequence 42077, A
15	217	4.1	1255	US-09-080-897-4	Sequence 4, Appl
16	217	4.1	1255	US-08-899-595-1	Sequence 1, Appl
17	217	4.1	1255	US-09-323-735-4	Sequence 4, Appl
18	214.5	4.1	506	US-09-949-016-11282	Sequence 11282, A
19	209.5	4.0	503	US-09-598-287A-2	Sequence 2, Appl
20	208	3.9	507	US-09-598-287A-24	Sequence 24, Appl
21	204.5	3.9	2321	US-09-230-652-2	Sequence 2, Appl
22	201	3.8	581	US-09-949-016-9978	Sequence 9978, Ap
23	199	3.8	802	US-09-823-240A-2	Sequence 2, Appl
24	198	3.8	709	US-09-949-016-6809	Sequence 6809, Ap
25	198	3.8	728	US-09-949-016-7213	Sequence 7213, Ap
26	195	3.7	62	US-08-687-702-18	Sequence 18, Appl
27	195	3.7	1312	US-09-345-882-29	Sequence 29, Appl

28	195	3.7	1315	3	US-08-899-595-3	Sequence 3, Appl
29	194.5	3.7	559	4	US-10-116-370-2	Sequence 2, Appl
30	194.5	3.7	567	4	US-09-949-016-10952	Sequence 10952, A
31	194	3.7	1248	2	US-09-080-897-2	Sequence 2, Appl
32	194	3.7	1248	3	US-09-323-735-2	Sequence 2, Appl
33	192.5	3.7	3248	1	US-08-353-700-1	Sequence 1, Appl
34	192.5	3.7	3248	5	PCT-US95-16216-1	Sequence 1, Appl
35	189.5	3.6	1572	2	US-08-290-731C-5	Sequence 5, Appl
36	189.5	3.6	1596	3	US-09-356-952-3	Sequence 3, Appl
37	189	3.6	1319	2	US-08-290-731C-2	Sequence 2, Appl
38	189	3.6	1336	2	US-08-290-731C-6	Sequence 6, Appl
39	189	3.6	2471	1	US-08-185-432-16	Sequence 16, Appl
40	189	3.6	2471	1	US-08-083-550A-19	Sequence 19, Appl
41	189	3.6	2471	3	US-08-532-384-19	Sequence 19, Appl
42	189	3.6	2471	4	US-08-899-232-1	Sequence 1, Appl
43	189	3.6	2471	4	US-09-121-452-1	Sequence 1, Appl
44	187.5	3.6	1187	3	US-08-664-962B-8	Sequence 8, Appl
45	187.5	3.6	1187	3	US-09-311-743-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-09-023-905A-12  
; Sequence 12, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023.905A  
; CURRENT FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-023-905A-12

Query Match	99.6%	Score 5253	DB 4	Length 1006	
Best Local Similarity	99.7%	Pred. No. 0			
Matches 1003	Conservative	0	Mismatches	3	Indels 0
Gaps	0				
Qy	1	MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAIIEALDVRMLVKMKKSVA	60		
Db	1	MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAIIEALDVRMLVKMKKSVA	60		
Qy	61	INSSGLAHVNEEQVTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKLTALFKNLIQNN	120		
Db	61	INSSGLAHVNEEQVTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKLTALFKNLIQNN	120		
Qy	121	NIISFPDLSLLKGLDKGVKGLKKPDKAWKDYETKITKIEKKEKHAHLHGMIRTEISG	180		
Db	121	NIISFPDLSLLKGLDKGVKGLKKPDKAWKDYETKITKIEKKEKHAHLHGMIRTEISG	180		
Qy	181	AEIAEEMEKERRFFOLQNCVCEYLLKVKNEIKIKGVDLLQNLIKYFHAQCNFFODGLKAVES	240		
Db	181	AEIAEEMEKERRFFOLQNCVCEYLLKVKNEIKIKGVDLLQNLIKYFHAQCNFFODGLKAVES	240		
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLQLRDLILKALQVEOKDSQIRQSTAYSJLHQPG	300		
Db	241	LKPSIETLSTDLHTIKQAQDEERQLQLRDLILKALQVEOKDSQIRQSTAYSJLHQPG	300		

Same as W0

```
Qy 301 NKEHGTNRGSLYKSGDGRKWKVKCSVKNQGFLLTISHGTANRPPAKLNLLTCQVKTNP 360
Db 301 NKEHGTNRGSLYKSGDGRKWKVKCSVKNQGFLLTISHGTANRPPAKLNLLTCQVKTNP 360
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBELT 420
Db 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBELT 420
Qy 421 KEIISEVQRMGTNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGVHYSFPMQSLTLDV 480
Db 421 KEIISEVQRMGTNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGVHYSFPMQSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGRSDMNAKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGRSDMNAKDYITAKYIERRYARKKH 540
Qy 541 ADNAAKHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHALHLAVRSVDRTS 600
Db 541 ADNAAKHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHALHLAVRSVDRTS 600
Qy 601 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHECEELLTQALSGRFSNVHVEYEWRLHEDLDESDDMDEKLOPSENRRDRP 720
Db 661 AKRLKHECEELLTQALSGRFSNVHVEYEWRLHEDLDESDDMDEKLOPSENRRDRP 720
Qy 721 ISFYQLGNSQLQSNVSLARDAANLAKEKORAFMPSILQNETYGALLSGSPPPAPQAPPS 780
Db 721 ISFYQLGNSQLQSNVSLARDAANLAKEKORAFMPSILQNETYGALLSGSPPPAPQAPPS 780
Qy 781 TTSAPPLPRNVGKVQTAASSANTLWKTNSVSDGSGRQSSDDPAVHPPPLPLAVSTN 840
Db 781 TTSAPPLPRNVGKVQTAASSANTLWKTNSVSDGSGRQSSDDPAVHPPPLPLAVSTN 840
Qy 841 PLTTPPPPPVAKTSPVMEALQSPKAPPGISQIRPPPLPQPPSRLPQKPKAPCTDKST 900
Db 841 PLTTPPPPPVAKTSPVMEALQSPKAPPGISQIRPPPLPQPPSRLPQKPKAPCTDKST 900
Qy 901 PLTNKQGRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLPKRVKALYNCAVNDP 960
Db 901 PLTNKQGRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLPKRVKALYNCAVNDP 960
Qy 961 DELTFSEGDIIVDGEEDQEWIIGHIDGDPGRKGAPVSVFVFIAD 1006
Db 961 DELTFSEGDIIVDGEEDQEWIIGHIDGDPGRKGAPVSVFVFIAD 1006
```

## RESULT 2

```
US-09-023-905A-7
; Sequence 7, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 982
; TYPE: PRT
; ORGANISM: Danio rerio
```

```
US-09-023-905A-7
Query Match 73.5%; Score 3874; DB 4; Length 982;
Best Local Similarity 74.1%; Pred. No. 8.2e-280;
Matches 755; Conservative 92; Mismatches 122; Indels 50; Gaps 13;
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRTNTVAATIEEALDVRMVLVKMKKS VKA 60
Db 1 MPDQITVAEFTVETNEDYKAPTASNTFTTRMTHCENTVSALEALDVRMVLVKMKKS VKA 60
Qy 61 INSSGLAHVNEEYOYTALEKFGCNCVCRDDPDIGSAPFLKPSVFTKELTALFKNLQNMN 120
Db 61 IYASGLAHVNEEYOYTALEKFGCNCVCRDDPDIGSAPFLKPSVFTKELTALFKNLQNMN 120
Qy 121 NIISFPLDLSLLKGDLLKGVKGDLLKVPDKAWKDYETKTKIEKKEKHAUKLHGMTRTISG 180
Db 121 NIITFPLDLSLLKGDLLKGVKGDLLKVPDKAWKDYETKTKIEKKEKHAUKLHGMTRTISG 180
Qy 181 AEIAEEMEKEKRRFPQLOMCEVLLKVNIEIKKGVDDLQNLIKYPHACNPFQDGLKAVES 240
Db 181 AEIAEEMEKEKRRFPQLOMCEVLLKVNIEIKKGVDDLQNLIKYPHACNPFQDGLKAVDN 240
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKEDSQIRQSTAYSLSHQPOG 300
Db 241 LKPSIEKLTDLHSIKQVQDEERRQLTQLRDVLKTLQVEQKEDSQVRSATYSLSHQPOG 300
Qy 301 NKEHGTNRGSLYKSGDGRKWKVKCSVKNQGFLLTISHGTANRPPAKLNLLTCQVKTNP 360
Db 301 NKEHGTNRGSLYKSGDGRKWKVKCSVKNQGFLLTISHGTANRPPAKLNLLTCQVKTNP 360
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBELT 420
Db 361 EEKCSFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBELT 420
Qy 421 KEIISEVQRMGTNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGVHYSFPMQSLTLDV 480
Db 421 KAILGEVKRMAGNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGVHYSFPMQSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGRSDMNAKDYITAKYIERRYARKKH 540
Db 481 LSTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGRSDMNAKDYITAKYIERRYARKKH 540
Qy 541 ADNAAKHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHALHLAVRSVDRTS 600
Db 541 PDALSKLHCLDCAVKARDIFSLIQVYAEVDLMPEPIPLANGHEQGETALHLAVLVDRTS 600
Qy 601 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIIDFLLTQNSLNDLKQTKAGSTALHYCCLTDNSECLKLLRGKASIDIANEAGETPLDI 660
Qy 661 AKRLKHECEELLTQALSGRFSNVHVEYEWRLHEDLDESDDMDEKLOPSENRRDRP 720
Db 661 ARRLKHLQCEELLQALAGFNAHVHVEYEWRLQHEDEDEDEDEKSSP--HRRDRP 718
Qy 721 ISFYQLGNSQLQSNVSLARDAANLAKEKORAFMPSILQNETYGALLSGSPPPAPQAPPS 780
Db 719 ISCTPGNSQLQSPASLSRDRDLVDKQKQFVFNLVNNETYGTINTSP-----VSL 772
Qy 781 TTSAPPLPRNVGKVQTAASSANTL-----WKTNSVSDGSGRQSSDDPAVHPPPLPLR 835
Db 773 SSSAPPLPRNL--VQPSALAGLTQSGSPGWKPGSLDLSG--RQRSSDDPPNMHPAPPLR 828
Qy 836 VTSTNPLTPT-PPPPVAKTSPVMEALQSPK--APPGISQIRPPPLPQPPSRLPQKPK 892
Db 829 VTSTSLMPSGAAPPLAKATGMETNMNQPXGGQPPGQINR-----871
Qy 893 APGTD---KSTPLTNKQGRGPDVLSATEAL--GPLSNAMVLQPPAMPKRSQATKLPK 947
Db 872 ATSAKQKFSKTLMRS-----GSIERPAKEVPGQPNQTTGQTLATHMPK---TYLAPK 923
Qy 948 RVKALYNCAVNDPDELTFSEGDIIVDGEEDQEWIIGHIDGDPGRKGAPVSVFVFIAD 1006
Db 924 RVKAMYNCAVNDPDELTFSEGELIIVDGEEDQEWLGHIEGPMRRGAPVTVQFIMD 982
```

```

RESULT 3
US-09-023-905A-10
; Sequence 10, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Danio rerio
US-09-023-905A-10

Query Match      59.8%; Score 3155; DB 4; Length 903;
Best Local Similarity 65.3%; Pred. No. 2.6e-226;
Matches 621; Conservative 98; Mismatches 174; Indels 58; Gaps 10;

Qy      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMVLKMKSVKA 60
Db      1 MPDQITVTFVDITHEDYKAPTISVFCETMAHCRNTVAALBEALDVRMVLKMKSVKA 60

Qy      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDLSAFKFSVFTKELTALFKNIQNN 120
Db      61 INSSGQTHVNEEQYTOAERTDNTVYKDDPEMSNYFLTFAFTKELTALFKNIQNN 120

Qy      121 NIISPLDLSLLKGDLLKGVKGDLLKPPDKAWKDYETKIKIEKEKEHAKLHGMITEISG 180
Db      121 NIITFPLDLSLLKGDLLKGVKGDLLKPPDKAWKDYETKLSKIEKEKEHAKLHGMITEISG 180

Qy      181 AEIAEMEKERRFQLOMCEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Db      181 GEIAEMEKERRFQLOMCEYLLKVEIKKGVLDLHNLKLYFHAQCNFFODGLKAVDN 240

Qy      241 LKPSIETLSTDLHTTIKQADEERRQLIQRLDILKSALQVEQKESQIRQSTAYSLSHQPG 300
Db      241 LKPFMEKATDILTANKQTOAERKQLQLKETLSALQSECKEDAQSKQACYSLSHQLOQ 300

Qy      301 NKEHGTENGSLYKSDGIRKWQKRSVKGFLTISHGTANRPPAKNLLTCQVKNP 360
Db      301 NKAHGTESGMLLKSEGLRWQKRSVKNGLLTISHGTNPANPNLLTCQVKNP 360

Qy      361 BEKKCFDLISHDRTYHFOADESOECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVOEIT 420
Db      361 DEKKCFDLISHDRTYHFOADESOECQVWVSLQNSKEEALNNAFKDDQNGENNIVRELT 420

Qy      421 KEIISVQRMGTNDVCCGADPTWLSNLGILTCIECSGIRHRELVHYSMPQSLTLDV 480
Db      421 KAIVGEVKWMSNDVCCGASNPWTLSNLGLVLCIECSGIRHRELVHYSRIQSLTLDL 480

Qy      481 LCTSELLAKNIGNAGFNEIMECCLPADSVPKPGSDMNARKDYITAKYIERRYARKKH 540
Db      481 LCTSELLANSVGNAAFNEIMEAKLSSE-IPKPYFSSDMQVRKDFITAKYTEKFAQKY 539

Qy      541 ADNAKHLHSLCAVKTDRDIFGLIQAYADGVOLTEKIPLANGHEPDETALHLAVRSVDRFS 600
Db      540 ADNAARLHALCAVAKSRDIFSLIQVABGLDMETINQPNQHEPGETSLHLAVRMVRNS 599

Qy      601 LHIVDFLVQNSGNDLKQTAGSTALHYCCCLTDNSECMLLRGKASVITNDAGETALDGL 659
Db      601 LHIVDFLVQNSGNDLKQTAGSTALHYCCCLTDNSECMLLRGKASVITNDAGETALDGL 659

Qy      661 AKRLKHECEBELLTQALSGRNSHVHVEYWRLLHEDDESDDDNDEKLQPSENREDRP 720
Db      660 AORLKHSKCEBELLTQAGTKFNHVHVEYWRLLHEDDESEDEDEK--PPIRREBRP 717

Qy      721 ISFYQLGNSQLQSNVSLARDAANLAKKQAFMPSILQNETYTGALLSGSPPAQPAAPS 780
Db      718 ISCVPGSGPMWPNMSALARDVANVNNKQAFIPSMNMNETYGTMLDPNSPL--GLPG 775

Qy      781 TTSAPPLPPRNKVKQVQTASSANTLWKTNSVSDGSRQRSSSDPPPAVHPPLPLAVTSTN 840
Db      776 VPGIPLPPRFLGRG-----W---SPMENIGRQSCSDPA---NPQTEQNNSVY 820

Qy      841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPPPPPPRLPQKPKPAGTDXST 900
Db      821 VLPAPPPPPA-----PKRPPPPDKASLLPAAATAP-----PAP-----SA 857

Qy      901 PLTNKGQPRGPDVLSATEALGPLSNAMVLOPPAPMPRKSOATKLPKRVKA 951
Db      858 PLLTPPAPLRPAP-----VPPAPVMPPTSSLTQVKSLLSKA 893

RESULT 4
US-09-023-905A-2
; Sequence 2, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Bos sp.
US-09-023-905A-2

Query Match      57.1%; Score 3013.5; DB 4; Length 1129;
Best Local Similarity 55.3%; Pred. No. 1.3e-215;
Matches 627; Conservative 129; Mismatches 225; Indels 153; Gaps 23;

Qy      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMVLKMKSVKA 60
Db      21 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMVLKMKSVKA 80

Qy      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDLSAFKFSVFTKELTALFKNIQNN 120
Db      81 IYNSGQHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFVKFSTLTKELTLLKLLQGLS 140

Qy      121 NIISPLDLSLLKGDLLKGVKGDLLKPPDKAWKDYETKIKIEKEKEHAKLHGMITEISG 180
Db      141 HNVITLDSLLKGDLLKGVKGDLLKPPDKAWKDYETKIKIEKEKEHAKLHGMITEISG 200

Qy      181 AEIAEMEKERRFQLOMCEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Db      201 AEIAEMEKERRFQLOMCEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKATK 260

Qy      241 LKPSIETLSTDLHTTIKQADEERRQLIQRLDILKSALQVEQKESQIRQSTAYSLSHQ 297
Db      261 LKQYIEKLAADLYNIKQTDDEKQLTALRDLIKSLQDQKESRRDSQSG--GYSMHQ 319
```

```
Qy 298 POGNKEHCTERNGLYKSGDGIKRWQKRCVKNGELTISHGTANRPPAKNLNLTTCOVK 357
Db 320 LQGNKEYSEKGYLLKKSGDGIKRWQKRCVKNGELTISHGTANRPPAKNLNLTTCOVK 379
Qy 358 TNPEKKCFDLISHDRTHYFQAEDEQECQIWMVLSQNSKEEALNNAFQDGDNTGNNIVQ 417
Db 380 PNAEDKKSFDLISHNRTYHFAQEDQDYVAMISVLTSNKEEALNNAFQDGDNTGNNIVQ 438
Qy 418 ELTVEIITSEVQRMGTNDVCCGAPDPTWLSNLGILTCIECSGHRBLGVHSPMQLT 477
Db 439 ELTKAIIEDVQRLPGNDVCCGAPDPTWLSNLGILTCIECSGHRBLGVHSPMQLT 498
Qy 478 LDVLTGSELLAKNIGNAGNEIMECCLPKPAEDSVKPNQSGDMNAPKDYITAKYIERRVAR 537
Db 499 LDKLTGSELLAKNIGNAGNEIMECCLPKPAEDSVKPNQSGDMNAPKDYITAKYIERRVAR 557
Qy 538 KKHADNAKLSLCEAVTRIDIFGLQAYADGVDLTEKIPLANGHEPDETAHLHLAVRSD 597
Db 558 KTCSSSAKLNELLEAIKSRDOLLALIQVYAEVGLWEPL-LEPCQELGETALHLAVRTAD 616
Qy 598 RTSLHIVDFVQNSGNDKQTKGKSTALHYCCLTDAECLLLRKGKASIEIANESGETP 657
Db 617 QTSLHLVDFVQNSGNDKQTKGKSTALHYCCLTDAECLLLRKGKASIEIANESGETP 676
Qy 658 LDIKRAKHECEELLTQALSGRFSNHHVVEYEWRLHEDLDESDDDMDKLOPSENRE 717
Db 677 LDIKRAKHECEELLTQALSGRFSNHHVVEYEWRLHEDLDESDDDMDKLOPSENRE 736
Qy 718 DRPISFYQLGNSQNSVLSARDANALAKQKRAFMPISILQNETYGA-----LLSGSP 771
Db 737 PRPOSFCHSSSISPD---KLSLPGFSTPRDKQL-----SYGAFTNQIFVSTSTD 784
Qy 772 PPAQAPASTTSAPPLPRNVGK-----VQTASSANTLWKTNSVSVGSGSRK 819
Db 785 SPTSPIA---EAPPLPRNATKGGPPGSTPLPLSTQSSGSLTKKRSPPPPPG-HKR 839
Qy 820 SSSDPPAVHPLPLPL--VTSNPLTTPPPVAKTPSPVMEALSQ-----862
Db 840 TSLDPPSPPLGPPNKGAVPWNQDVGPSSS---SKTINKFELSGQSSGTSKAKTALVPRV 896
Qy 863 ---PSKPA-----PPGI---SQI-----RPPQ---LPPQ-----883
Db 897 LPKLPQKVALRKTSLSHLSDKANVPPEIQKSSQLTELQKPPGDLPPKPTLAPKP 956
Qy 884 ---PSRLPQK-----KPAAGTDSKSTPLTNKGQPR 909
Db 957 PIGDLPKPGELPPKPPQLGDLPPKQVADLPPKQVADLPPKQVADLPPKQVADLPPK 1016
Qy 910 G-----PVDLS-----ATEALGPLSNAMVLOPPAPMPKSKOATKLPKRVKAL 952
Db 1017 AQPPELTPKSHPADLSPNVKQASEDNDVFTPLP-ETPVPPLPKINTGSKVRVRKTI 1075
Qy 953 YNCVADNPDELTFSEGVIIIVDGEDEQEWIGHIDGDPGRKGAFPVSVFHTAD 1006
Db 1076 YDCQADNDELTFMEGEVIVTGEDEQEWIGHIEGQPERKGVFPVSVFHTAD 1129
```

## RESULT 5

```
US-09-023-905A-4
; Sequence 4, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
```

```
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Danio rerio
; US-09-023-905A-4
```

```
Query Match 56.0%; Score 2954.5; DB 4; Length 1151;
Best Local Similarity 54.1%; Pred. No. 3.2e-211;
Matches 622; Conservative 126; Mismatches 239; Indels 163; Gaps 24;

Qy 1 MPDQISVSFVETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSVKA 60
Db 21 MPDQISVSFVETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSVKA 80
Qy 61 INSSGLAHVNEREOYTOALEKFGNGCVCRDDPDGLGSFLKPSVFTKELTALFKNLIONMN 120
Db 81 IYNSGQEHVQNEENYGOALDKFGNSFISRDNDSLTGTAIFKPSGLIKELAAKLLKQLSLS 140
Qy 121 NIIFFPLDLSLLKGLKGVKDLKKPFDKAWKDYETKITKIEKKEKHAHGMIRTEISG 180
Db 141 HNVFTLDSLLKGLKGVKDLKKPFDKAWKDYETKITKIEKKEKHAHGMIRTEITG 200
Qy 181 AEIAEAEKERRFPQLOMCEYLLKWEIKKGVLDLONLIKYFHAQCNFPQDGLKAVES 240
Db 201 AEIAEAEKERRFPQLOMCEYLLKWEIKKGVLDLONLIKYFHAQCNFPQDGLKAVES 260
Qy 241 LKPSIETLSLTHTKQAQDEERQLOLRDLKASALQVEOKEDSQRSTAYSLSHPOQG 300
Db 261 LKQIETLADLYNLIKQDDEKQKQALRDLIKSSQDQKEDSQSKS-GYSHMLQOG 319
Qy 301 NKEHGTBRNSGLYKSGDGIKRWQKRCVKNGELTISHGTANRPPAKNLNLTTCOVKTNP 360
Db 320 NKEHGTBRNSGLYKSGDGIKRWQKRCVKNGELTISHGTANRPPAKNLNLTTCOVKTNP 379
Qy 361 BEKKCFDLISHDRTHYFQAEDEQECQIWMVLSQNSKEEALNNAFQDGDNTGNNIVQELT 420
Db 380 BEKKCFDLISHDRTHYFQAEDEQECQIWMVLSQNSKEEALNNAFQDGDNTGNNIVQELT 438
Qy 421 KEIITSEVQRMGTNDVCCGAPDPTWLSNLGILTCIECSGHRBLGVHSPMQLTLDV 480
Db 439 KAITEDVLRIPGNEVCCCGVPEKWLSTNLGILTCIECSGHRBLGVHSPMQLTLDV 498
Qy 481 LGTSELLAKNIGNAGNEIMECCLPKPAEDSVKPNQSGDMNAPKDYITAKYIERRVAR 540
Db 499 LGTSELLAKNIGNAGNEIMECCLPKPAEDSVKPNQSGDMNAPKDYITAKYIERRVAR 557
Qy 541 ADNAKLSLCEAVTRIDIFGLQAYADGVDLTEKIPLANGHEPDETAHLHLAVRSDRVS 600
Db 558 TTATARQCDLYEAVTRIDIFGLQAYADGVDLTEKIPLANGHEPDETAHLHLAVRSDRVS 616
Qy 601 LHIYDELVQNSGNDKQTKGKSTALHYCCLTDAECLLLRKGKASIEIANESGETPDI 660
Db 617 LHIYDELVQNSGNDKQTKGKSTALHYCCLTDAECLLLRKGKASIEIANESGETPDI 676
Qy 661 AKRLKHECEELLTQALSGRFSNHHVVEYEWRLHEDLDESDDDMDKLOPSENREDRP 720
Db 677 AKRLKHECEELLTQALSGRFSNHHVVEYEWRLHEDLDESDDDMDKLOPSENREDRP 736
Qy 721 ISFYQLGNSQNSVLSARDANALAKQKRAFMPISILQNETYGA-----AFM-----754
Db 737 QSF-----CHSSSVS--PQEKLTLPGLYGHDRKQRLSYGAFANPVPVSTSTETPASPVSE 788
Qy 755 -PSILQNETYCALLSGSP-----PPAQAAPASTTSAPPLP---788
Db 789 GPTTAKTPAKAPSCGPTSLPLGSSQSSGSLTKKRSPPPPGKRTSHDPPSPVLQ 848
Qy 789 -PRNVGKVQTTASSANTLWKTNSVSVGSGSRQSSSD-----PPAVHPLP-----831
```



Db 849 GPQSGSESTPPSANRTSPAN--KEGICQOOSTTSMNTKATFGPRVLPKLPQKVALRK1 906  
QY 832 -----PPLRVTSNPLT-PTPPPVAKTSPVNEALSQPSK 865  
Db 907 DTIHLPSVDKSGPDLVQKPPQAQDAPPTASDT--ITRTEPPP--KIPQVAERSQPPDV 962  
QY 866 PAPGISOIRPPP-----LPPQ-----PSRLPQKXPACTDKSTPLTNKGQ-----PRGPV 912  
Db 963 POKPHISDLPPKPOLSDLPKPOLSDLPKPOLSDLPKPOLSDLPKPOLSDLPKPOLSDLPK 1021  
QY 913 DLSATEA-----IGPL-----SNAMVLQPPAMP 937  
Db 1022 VCSASEATQRTQSTETSPKQLTETQSFQSEELSPQASEDTNGAPAGALEMVPNPR 1081  
QY 938 K-SQATKUPKRVKALYCNADNPDLTFSEGDIIVDGEEDQEWIIGHIDGPKRGAP 996  
Db 1082 KINTVAKNKAARVKTYYDCQADNDDELTFVEGEVIIVTGEEDQEWIIGHIEQPERKGVF 1141  
QY 997 PVSFVHFIAD 1006  
Db 1142 PMSFVHILSD 1151

## RESULT 6

US-09-023-905A-36

; Sequence 36, Application US/09023905A

; Patent No. 6475778

; GENERAL INFORMATION:

; APPLICANT: Roberts, Thomas M.

; APPLICANT: King, Frederick J.

; APPLICANT: Harris, David F.

; APPLICANT: Hu, Erding

; APPLICANT: Spiegelman, Bruce

; APPLICANT: Chan, Joanne

; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses

; FILE REFERENCE: DFN-021

; CURRENT APPLICATION NUMBER: US/09/023,905A

; CURRENT FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/038,191

; PRIOR FILING DATE: 1997-02-14

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 36

; LENGTH: 307

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: consensus sequence

US-09-023-905A-36

Query Match 26.3%; Score 1387; DB 4; Length 307;  
Best Local Similarity 88.9%; Pred. No. 2.9e-95;  
Matches 273; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMLVYKMKSVKA 60  
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRLAHCRNTVAABEALDVRMLVYKMKSVKA 60  
QY 61 INSSGLAHVENEQYTOALEKFGNVCRRDDPLGSAFLKFSVFTKLTALFKNLIQNN 120  
Db 61 IYSSGQAHVENEQYTOALEKFGNVCRRDDPLGSAFLKFSVFTKLTALFKNLIQNN 120  
QY 121 NIISFPLSLGDLKGVKGDLPKDPKAWKDYETKTKIEKKEKHAHGMIRTEISG 180  
Db 121 NIISFPLSLGDLKGVKGDLPKDPKAWKDYETKTKIEKKEKHAHGMIRTEISG 180  
QY 181 ABIAEEMEKERRFFLOMCEYLLKYNKIKKGVLLQNLIKYFHAQCNFFQDGLKAVES 240  
Db 181 ABIAEEMEKERRFFLOMCEYLLKYNKIKKGVLLQNLIKYFHAQCNFFQDGLKAVDN 240  
QY 241 LKPSIETISTDLHTTKQAQDEERRQLIQLRDILKSALQVEQKEDSQIRQSTAYSILHQPG 300  
Db 241 LKPSIETISTDLHTTKQAQDEERRQLIQLRDILKSALQVEQKEDSQIRQSTAYSILHQPG 300

Db 241 LKPSIEKATDLHNIKQTDERRKQLTQLRDILKSALQVEQKEDSQIRQSTAYSILHQPG 300  
QY 301 NKEHGTG 307  
Db 301 NKEHGTG 307

## RESULT 7

US-09-023-905A-30

; Sequence 30, Application US/09023905A

; Patent No. 6475778

; GENERAL INFORMATION:

; APPLICANT: Roberts, Thomas M.

; APPLICANT: King, Frederick J.

; APPLICANT: Harris, David F.

; APPLICANT: Hu, Erding

; APPLICANT: Spiegelman, Bruce

; APPLICANT: Chan, Joanne

; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses

; FILE REFERENCE: DFN-021

; CURRENT APPLICATION NUMBER: US/09/023,905A

; CURRENT FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/038,191

; PRIOR FILING DATE: 1997-02-14

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 30

; LENGTH: 134

; TYPE: PRT

; ORGANISM: Bos sp.

US-09-023-905A-30

Query Match 9.2%; Score 484.5; DB 4; Length 134;  
Best Local Similarity 67.4%; Pred. No. 1.5e-28;  
Matches 91; Conservative 19; Mismatches 24; Indels 1; Gaps 1;  
QY 421 KEIISVQRMVTGNDVCCDGPADPTWLTNLTGILTCIECSGTHRELVGHVYSPQSLTLDV 480  
Db 1 KAIIEDVQRLPGNDVCCDGPADPTWLTNLTGILTCIECSGTHRELVGHVYSPQSLTLDV 480  
QY 481 LGTSELLAKNIGNAGNEIMECCLPADSPKPNPGSDMNAKDYITAKYIERRVARKKH 540  
Db 61 LGTSELLAKNIGNAGNEIMECCLPADSPKPNPGSDMNAKDYITAKYIERRVARKKH 540  
QY 541 ADNAAKHSLCEAVK 555  
Db 120 SSSAKLNELLEAK 134

## RESULT 8

US-09-270-767-33258

; Sequence 33258, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 33258

; LENGTH: 166

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-33258

Query Match 8.6%; Score 453; DB 4; Length 166;  
Best Local Similarity 57.1%; Pred. No. 4.5e-26;  
Matches 84; Conservative 30; Mismatches 33; Indels 0; Gaps 0;  
QY 2 PDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMLVYKMKSVKA 61

Db 20 PSLIAVSEFVEETRSYSPITTSFASRMPCDRTHTIGVLEERLEFDRGLTKLKKAVKAI 79  
Qy 62 NSSGLAHVENEQYTOALEKFGGNCVCDDPDLGSAFKFSVFTKELTALFKNLIONMNN 121  
Db 80 HNSGNTHVDNMFVVRALERLGGKVIQDEPDIGAAFLKFSVVTKELSALMKTLMOINN 139  
Qy 122 IISFPLDSLKGLKGVKGLKKPFDK 148  
Db 140 IVMPFVDSMLKSELKRGVAGDMKRPFDK 166

RESULT 9  
US-09-828-303-19  
; Sequence 19, Application US/09828303  
; Patent No. 6677504  
; GENERAL INFORMATION:  
; APPLICANT: COSTA E SILVA, OSWALDO DA  
; APPLICANT: BOHNET, HANS J.  
; APPLICANT: VAN THIELEN, NOCHA  
; APPLICANT: CHEN, ROUTING  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND  
; FILE REFERENCE: METHODS OF USE IN PLANTS  
; CURRENT APPLICATION NUMBER: US/09/828,303  
; CURRENT FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/196,001  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Physcomitrella patens  
US-09-828-303-19

Query Match 7.9%; Score 419; DB 4; Length 605;  
Best Local Similarity 23.9%; Pred. No. 9.9e-23;  
Matches 141; Conservative 96; Mismatches 198; Indels 156; Gaps 19;

Qy 211 KKGVDLLQNLKYPHAQCNFFODGLKAVESLKPSIETLSTDLHTIKQAQDEERRLIQLR 270  
Db 10 KKKPFLEAVSGTMDAHLRYPKQGYELLHQBPIYHQVLT---YAQGRERANYEQAALA 66  
Qy 271 DILKSALQ-VEQKESQIR-----QSTAYSILHQ-----PQGNKEHGTERNGS 311  
Db 67 DRMQEYRQEVERESQSIQDFSSSGDGIQGVGRSSHKMIEAVMQSTPKGQIQ--TLKQGY 124  
Qy 312 LYKKSQDGIKVMQKESV-KNGFL-----TISHGTANRPPAKLNLTLTQV 356  
Db 125 LLKSTNLGDKWRKRRFFVLDSRGLMYRKYKQWKGKPTDEKNVAHT-----VNLITSTI 177  
Qy 357 KTNPEE---KKCFDILSHDRTHYFAQDEQECQIWMVSLQNSKEBALNNAFAGDDNTGEN 413  
Db 178 KIDASQSLRFRFCRIISPAKSYTLQAEINAIDRMWDMDKITGVISLLNNQI-SEQVDGED 236  
Qy 414 NIV-----QELTEIIESEVORMTGVNDVCCGAPDPTWLTSTNLGILTCESGIHREL 466  
Db 237 SDVSRGASDQSGHERPLDLVRKGVNDACADCGAADPDWASLNLGILLCTECSGVHRNM 296  
Qy 467 GVHYSFPMOSLTLD--VLGTSELLAKNIGNAGFNEIMECCLPAEDS-----510  
Db 297 SVQISKVRSLLTUDVKWVPSVMSYFQSGVNSVANSIWEELNPKSSESSERNVNDGQS 356  
Qy 511 -----VKPNFGSDMNARKYITAKYIERRYARKKHADN--AAKLHSLCEAVKTRDIFG 561  
Db 357 GVLASARPRPRDPIPIKERFINAKYVEKFKVQKLKVDSPRGVSTRQIWDVAVQNKVQL 416  
Qy 562 LLQAY--AD-----GVDLTEKIPLAN-----580  
Db 417 ALRLLITADANANTTFEQVMGGTEGSSWSPLASAGALLRKNLSASQSGRRNRWSVPSLL 476  
Qy 581 -----GHEPDE-----TALHLAVRSVDRSTLSLHIVDFLVQN 610

Db 477 SSPDDPGSRGALSPVSRSPDAAGSGGIDEKDLRGCSLLHVACQIGD---ISLIELLQY 533  
Qy 611 SGNLDKQTKGSTRALHYCCCLTDNAECIKLLRGKASIEIANESGETPLDIA 661  
Db 534 GAQINCVDLTGRPLHLHCVLCGNNSCAKLLLTTRGAKAGAVDKGKTPLECA 584

RESULT 10  
US-09-270-767-42748  
; Sequence 42748, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 42748  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-42748

Query Match 6.6%; Score 347.5; DB 4; Length 225;  
Best Local Similarity 41.0%; Pred. No. 5.1e-18;  
Matches 87; Conservative 31; Mismatches 73; Indels 21; Gaps 6;

Qy 517 SDMNARKDYITAKYIERRYARKKHADNAAKLHSLCEAVKTRDIFGLLOAVADGVDLTEKI 576  
Db 6 SSMEERYDFIRAKYAKRYVMRTCSDDNDLRCDLEQAVVADMSQLQVWAGADLTCCCL 65  
Qy 577 PLANGHEPDETALHLAVRSVDRSTLSLHIVDFLVQN-----SGNLDKQTKGKSTA 624  
Db 66 P---SSDAGETALHLAVLREMGSTLHIVDFLIQNMPPKGLNKPATNPAGLLD-VTGG-NTA 120  
Qy 625 LHYCCLTDNAECLLARGKASIEIANESGETPLDIAKRLKHEHCELLTQALSGRFNSH 684  
Db 121 LHLCALHRRRCMKLLRSGADYELKNSQNTALDIAKEMGHNSCRELIECAIKREKSAF 180  
Qy 685 VHVEYWRLLHED--LDESDDD--MDEKLOPS 712  
Db 181 DHINTDNLNPNEDGSTDFSDDETVIDERSSSS 212

RESULT 11  
US-09-949-016-6235  
; Sequence 6235, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6235  
; LENGTH: 802  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6235

Query Match 5.3%; Score 278; DB 4; Length 802;

Best Local Similarity	20.1%;	Pred. No. 4.7e-12;
Matches	Conservative 141;	Mismatches 333; Indels 318; Gaps 47;
QY	42	EALDVRMVLVYKMKSVKAINSSGLAHVENEQYTOALEKFG-----GNCVCRD 90
Db	20	ERLKCVEQLERTNKFKIDVIKDGNALLSAMRNVSSAVQKFSQTLQSFPDFIGDTLTD 79
QY	91	DPDLGSAFLKFSVFTKELTALFKNLIQNMNNIISFPDLSLLKGDLLKGVGDLKLPFKAM 150
Db	80	EINIAESFKFAELNNEVNERMMVFNASLLIKPLENFRKEQI-GPTKRRKKKFKXDG 138
QY	151	KDYETKITK--TBKREKKEHAKLHGMIRTEISGAIEAEMEKEKRRFFQLQWCEVLLKYNE 207
Db	139	ERYFSLDRHLHLSSKKKE-----SQLEADL--QVDEKRNFPPESSLDVYVYQIE 187
QY	208	IKIKKGVDLLQNLKIYFHAQCNFFQDGL--KAVESLKPISIBTSLDHLTIKQAODEERRQ 265
Db	188	VQSSKKNFNEVPVLAFLHSL--FISNSLTVELTQDFLPYKQQLQLSLQTRNHFSSSTREE 245
QY	266	LIQRLDLKALQVEQKEDSQIROSTAYSLHQPOGNEKHGTER-NGSLY---KKSDGIRK 321
Db	246	MEELKKRMKEA-----POTCKLPQPTIEGVLVYTOEKRWALGIS- 283
QY	322	VWQKRKSVKNGFGLTIS-----HGTANRPPAKNLLTC-QVTKNPBEKK-CFDLISHDR 373
Db	284	-WKYICYQSEKTKLTWTMPWEQKPGAKGGLDUTLKVCVARKTESIDKRCFOIETNER 342
QY	374	--TYHQAQDEQECQIWMVSLQNSKEBALNNAPFGDDNTGNNIVQELTKETIISSEVQMT 431
Db	343	PGTITLQALSEANRRLWEAM-DGKEPIYHSPITKQEMELNEVGKPVKRCINIE-- 398
QY	432	GNDVCCDGGAPDTWLSTNIGILTICIECSGHIHRELGVHYSQMSQLTDLVLTSETLLAKN 491
Db	399	-----TKGIKT-----EGLRTVGSNIQVQKLL----- 421
QY	492	IGNAGFNEIMECCLPADSVKPNPGSDMNAKDYITAKYIERRYARKGHADNAAKLHSLC 551
Db	422	--NAFFDP--KC-----PGD-----VDFHNSD 439
QY	552	EAVKTRDIFGLQAYADGVDLTKIPLANGHEPDET-ALHLAVRSVDRSTLSLHVDLVQN 610
Db	440	WDIKT-----ITSSLKFYLRNLSEPVMTYRLHKELYSAK----- 474
QY	611	SGNLDKQTGKGTALH---YCLLTDNAECLKLLRGKASIEIANESGETPLDIAKRLKHE 667
Db	475	SDNLDYRLG---AIHSLVYKLPKREMLELLIRHLNV-----CE 512
QY	668	HCRE-LLTQALSGRFNSHVHVEYEWRLLEDLDESDDMDKLOPSENRRDRPISFYQL 726
Db	513	HSKENLTP-----SNMGVIFGPTLMRAQEDTVAAAMNLIKFO----- 549
QY	727	GSNQLQSNVSLARDAANLAKEQRAEMPSILQNETYCALLSGSPPPAQAAPTSTTAPP 786
Db	550	-----NIVVE-----ILLIEHFGKIVLG--PPESA-----APP 575
QY	787	LPPRNV---GKVQTASS---ANTLWKTNSVSDGGSRQR-----SSSDP--PAWH 828
Db	576	VPPRPVTAARHKPTITISKRLLRRTVFYTSLSDESEDEIQHTPNGTITTSIEPPKPPQH 635
QY	829	PPLPPLRVSTNPLTPTPPPVAKTPTPSVMEALSQPSKPAP-----PGISQIRP 876
Db	636	PKUPIQSGGTDGPKGSPSRPI-----LDKLEPCPEVDVGVKLVRLQDGGTKLTP 686
QY	877	P-----PLPPOPPSRLPQ---KKPAPGTDKSTPLTNKGQPRGVDLSATEALGPLSNAMVL 929
Db	687	KATNGMPGSGPTKTPSFHIKRPAP-----RPLAH--HKEGDAD-SFSKVRPPCKEPII 738
QY	930	QPPA---PMPKRSQATKLPKRVKALYNCVADNPDELTFSS 966
Db	739	RPPVRPDDPCRAATPQKPPPKP---DIVAGNAGEITSS 774

## RESULT 12

```

US-09-949-016-11166
; Sequence 11166, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11166
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11166

```

```

Query Match      5.0%; Score 262; DB 4; Length 750;
Best Local Similarity 20.1%; Pred. No. 6.7e-11;
Matches 194; Conservative 143; Mismatches 311; Indels 318; Gaps 47;

QY      75 YTOALSKFG-----GNCVCRRDDPLGSAFLKFSVFTKELTALFKNLQNMNII 123
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1 YSAVQKFSQTTQSFQDFIGDTLTDOETNIAESFKFEABLLNEVENRMMVHNAEDLL 60
      QY      124 SFPLDLLKGDGLKGVGDKLPKPKDAWKQYETIKT--IEKKKSHAKLHGIMRTISG 180
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
61 IKPLENFRKEQI-GFTKERKKKFKQGERFYSLLDRHLHLSKKKG-----SQLQE 110
      QY      181 AETAEMEKERRPFLQMCYELKLVNNEIKKIGVDDLQNLKIYFHAQCNFFQDGL--KAV 238
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
111 ADL--QVDKERHNFPESSLDVYQIOEVOESKKFNIVEPVLAFLHSL--FISNLSLTVELT 166
      QY      239 ESLKPSIETLSITDLHTIKQAOBERQQLQTLRDILKASALQVEQKDSQIQSTAYSLSHQ 298
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
167 QDPLPYKQQLQSLQWNRNHFSTRBEMELKIKMKEA-----P 205
      QY      299 QGNKEHGTER-NGSLY---KKSQGIKVKQKRCVKNGPLTIS-----HGTANRRPPAK 348
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
206 QTCKLPGQPTIEGYLYTQEKWALGIS--WKVYCYQYEKETKTLTMTPMQKPGAKQGPLD 263
      QY      349 LNLITC-QVKTNPBEKK-CFDLISHDR--TYHFQABDEQECQIWMSVLQNSKEALNNAF 404
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
264 LTLKYCVRRKTESIDKRCFDIETNRPGTITLQALSEANRLLWMEAM-DGKEPIYHSP 322
      QY      405 KGDNTGNNIVQELTKELISEVQRMGTNDVCCGAPDPFTWLSNLGILTCIECSIHR 464
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
323 TKQOEMELNEVGFKFVRKCNIIIE-----TKGIKT---EGLYR 357
      QY      465 ELGVHVSPMQSLTLDVLGTSELLAKNIGNAGFNEIMECCLPADSVKPNPGSDMNAKD 524
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
358 TVGSNTQVQKLL-----NAFDP-KC-----PGD-----380
      QY      525 YITAKYIERRYARKKHADNAAKLHSLCEAVKTRDIFGLLQAYAGVDLTETKPIPLANGHEP 584
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
381 -----VDFHNSWDIKT-----ITSSLUKPYLRNLSEP 407
      QY      585 DET-ALHLAVRSVDRTSLLHVPVQNSGLDKQTKGKSTALH---YCLTUDNAECLKLL 640
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
408 VMTYRLHKLVLVSAK-----SDNLDYRLG---AIHSLVYKLPKPKREMLELL 451
      QY      641 LRKKAIEIAENSGETPLDIAKELKHEHCEE-LLTQALSGRFSNVHVEYEWRLHLEDLD 699
      Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
452 IRLHNVN-----CEHSKENLMT-----SNMGVIFGPTLMRAQED 486
      QY      700 ESDDMDKLOPSENRRDRPISFYQLGNSQLQSNVSLARDANAIAKEKQRAFMPISILQ 759

```

7, -





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:11:14 ; Search time 30 Seconds  
(without alignments)  
3226.469 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 5273  
Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAPVSPVHFAD 1006

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_79.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5253	99.6	1006	2 T00050	hypothetical prote
2	3000.5	56.9	1147	2 T42627	ADP-ribosylation f
3	485.5	9.2	776	2 A96634	probable GCMA-comp
4	479.5	9.1	752	2 T26508	hypothetical prote
5	466	8.8	750	2 T48577	hypothetical prote
6	399	7.6	531	2 D86242	hypothetical prote
7	393.5	7.5	1087	2 T49496	hypothetical prote
8	335	6.4	870	2 T39731	csx2 protein - fis
9	319.5	6.1	903	2 T26743	hypothetical prote
10	319.5	6.1	951	2 T26738	hypothetical prote
11	319.5	6.1	975	2 T26737	hypothetical prote
12	309.5	5.9	923	2 T38398	hypothetical GTPas
13	278	5.3	802	2 H59434	oligophrenin 1, Rh
14	271	5.1	1210	2 C59431	centaurin, delta 2
15	267.5	5.1	814	2 F59430	GTPase regulator a
16	237.5	4.5	601	2 T11677	probable transcrip
17	229.5	4.4	1621	2 T15264	hypothetical prote
18	227.5	4.3	482	2 S69580	hypothetical prote
19	226	4.3	1201	2 G86441	unknown protein [i
20	223	4.2	1181	2 T30578	myosin IC - slime
21	221.5	4.2	334	2 A85067	hypothetical prote
22	221.5	4.2	334	2 S45306	notch 3 protein -
23	221	4.2	1544	2 E59431	phosphoinositide-b
24	217	4.1	1255	2 T31065	diaphanous protein
25	215.5	4.1	495	2 T26300	hypothetical prote
26	212.5	4.0	3421	1 W2B86	367K tegument prot
27	210.5	4.0	1044	2 T00342	hypothetical prote
28	210	4.0	1953	2 S63244	BNI1 protein - yea
29	208	3.9	403	2 S52796	prpL2 protein - hu

30	207.5	3.9	708	2 D96711	hypothetical prote
31	206	3.9	1837	2 T41023	probable nuclear p
32	204.5	3.9	369	2 T04947	hypothetical prote
33	204.5	3.9	2321	2 S78549	notch3 protein - h
34	202.5	3.8	1188	2 S49915	extensin-like prot
35	201.5	3.8	592	1 LLBY	actin-binding prot
36	199	3.8	315	2 A56561	35K proline-rich p
37	198.5	3.8	423	2 T23223	hypothetical prote
38	197	3.7	464	2 S22697	extensin - Volvox
39	196.5	3.7	352	2 S47006	zinc finger protei
40	196.5	3.7	907	2 E96636	hypothetical prote
41	196.5	3.7	1468	2 S11515	formin - mouse
42	196	3.7	818	2 A59433	KIAA0672 protein [
43	196	3.7	1282	2 J50120	glycoprotein A - m
44	195	3.7	1168	1 MWAXIC	myosin heavy chain
45	194	3.7	505	2 S72273	actin-depolymerizi

#### ALIGNMENTS

##### RESULT 1

T00050

hypothetical protein KIAA0400 - human

C:Species: Homo sapiens (man)

C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999

C:Accession: T00050

R:Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka, A.,

submitted to the EMBL Data Library, October 1997

A:Description: Prediction of the coding sequences of unidentified human genes. VIII. The

A:Reference number: Z14080

A:Accession: T00050

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1006 <ISH>

A:Cross-references: EMBL:AB007860; NID:d1175316; PIDN:BAA3696.1; PID:d1024577

A:Experimental source: brain, clone HGI091

C:Genetics:

A:Note: KIAA0400

Query Match 99.6%; Score 5253; DB 2; Length 1006;  
Best Local Similarity 99.7%; Pred. No. 2.9e-248;  
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRTNVAIAIEEALDVRMVLVKMKSKVKA	60
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRTNVAIAIEEALDVRMVLVKMKSKVKA	60
Qy	61	INSSGLAHVNEEQYTQALEKFGNCVCRRDDPDILGSAPLKPSVFTKELTALFKNLIQNN	120
Db	61	INSSGLAHVNEEQYTQALEKFGNCVCRRDDPDILGSAPLKPSVFTKELTALFKNLIQNN	120
Qy	121	NIISFPDLSLLKGLDKGVKGLKKPDKAWKDYTKTKIEKEKEHAHKLHGMTRTEISG	180
Db	121	NIISFPDLSLLKGLDKGVKGLKKPDKAWKDYTKTKIEKEKEHAHKLHGMTRTEISG	180
Qy	181	AEIAEEMEKEKRRFFQLOWCEYLLKKNVKKGGVLDLQNLKLYFHAQCNPFDQGLKAVES	240
Db	181	AEIAEEMEKEKRRFFQLOWCEYLLKKNVKKGGVLDLQNLKLYFHAQCNPFDQGLKAVES	240
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLIQRLDKLSALQVEQKESQIRQSTAYSLSHQPOG	300
Db	241	LKPSIETLSTDLHTIKQAQDEERQLIQRLDKLSALQVEQKESQIRQSTAYSLSHQPOG	300
Qy	301	NKEHGTNRNGLYKKSQIRKQWQKRCVKNQGLTISHGTANPPAKNLNLTQCVKTNP	360
Db	301	NKEHGTNRNGLYKKSQIRKQWQKRCVKNQGLTISHGTANPPAKNLNLTQCVKTNP	360
Qy	361	EEKKCPDLISHDRTYHFQADEQECQIWMVSLQNSKEALNNAFKGDDTGNENIVQELT	420
Db	361	EEKKCPDLISHDRTYHFQADEQECQIWMVSLQNSKEALNNAFKGDDTGNENIVQELT	420
Qy	421	KEIISVQRMGTGNDVCCDCCGAPDPTLWLTNLTGILTCIECSGIHRELGVHYSPMQSLTLDV	480

Db 421 KEIISVQRMGTNDVCCGAPDPTWLTSTNLGILTCIECSGIRHGLVHYSRMQSLTLDV 480  
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERYARKKH 540  
Db 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERYARKKH 540  
Qy 541 ADNAKLSHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDT 600  
Db 541 ADNAKLSHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDT 600  
Qy 601 LHIIVDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLLRGRKASIEIANESGETPLDI 660  
Db 601 LHIIVDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLLRGRKASIEIANESGETPLDI 660  
Qy 661 AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDMDKQLQSPENRRDRP 720  
Db 661 AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDMDKQLQSPENRRDRP 720  
Qy 721 ISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILQNETYGALLSGSPPPAQAAPS 780  
Db 721 ISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILQNETYGALLSGSPPPAQAAPS 780  
Qy 781 TTSAPPLPPRNKVGKVTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPLPLRVSTN 840  
Db 781 TTSAPPLPPRNKVGKVTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPLPLRVSTN 840  
Qy 841 PLTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRPLPQKPAFGTKST 900  
Db 841 PLTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRPLPQKPAFGTKST 900  
Qy 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNCVADNP 960  
Db 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNCVADNP 960  
Qy 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVSFVHFAD 1006  
Db 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVSFVHFAD 1006

RESULT 2  
T42627  
ADP-ribosylation factor-directed GTPase activating protein, isoform a - mouse  
N:Alternate names: ASAP1a protein  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T42627  
R:Brown, M.T.; Andrade, J.; Radhakrishna, H.; Donaldson, J.G.; Cooper, J.A.; Randazzo, P.  
Mol. Cell. Biol. 18, 7038-7051, 1998  
A:Title: ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates with  
A:Reference number: 222178; MUID:99038209; PMID:9819391  
A:Accession: T42627  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1147 <BRO>  
A:Cross-references: UNIPROT:Q9QWY8; EMBL:AF075461; NID:G4063613; PID:G4063614; PIDN:AAC9  
C:Genetics:  
A:Gene: Shag1  
C:Function:  
A:Description: interacts directly with ADP-ribosylation factors (Arf1, Arf5 and to a lesser  
A:Note: supposed to coordinate membrane remodeling events  
C:Keywords: membrane trafficking

Query Match 56.9%; Score 3000.5; DB 2; Length 1147;  
Best Local Similarity 53.9%; Pred. NO. 1.7e-138;  
Matches 621; Conservative 128; Mismatches 232; Indels 171; Gaps 19;  
Qy 1 MPDQISSEFVAETHEDYKAPTASSFTTACRNTVAIAEFLNDVDRMVLKMKSKVKA 60  
Db 21 MPDQISSEFVAETHEDYKAPTASSFTTACRNTVAIAEFLNDVDRMVLKMKSKVKA 80  
Qy 61 INSSGLAHVENEQYQTALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALFKNLIQNN 120  
Db 1136 VFPVSFVHILSD 1147

Db 81 IYNSQDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFAVKFSTLTKELTLLKNLQGLS 140  
Qy 121 NIISFPPLDSLLKGLKVGDLKKPPDKANKDVEYTKITKIEKKEKHAUKHGMTREISG 180  
Db 141 HNVIFTLDSLLKGLKVGDLKKPPDKANKDVEYTKITKIEKKEKHAUKHGMTREISG 200  
Qy 181 AETAEMEKEKRRFQLOMCEYLLKKGVDLLQNLIKYKHAOCNFPQDGLKAVES 240  
Db 201 AETAEMEKEKRRFQLOMCEYLLKKGVDLLQNLIKYKHAOCNFPQDGLKAVES 260  
Qy 241 LKPSIETSLDHTIKQAQDEERRQLIQLRDLKSLAQVEQKE-----DS 285  
Db 261 LKQVIEKLAADLYNIKQDDEEKQLTALADLIKSLQLDPPKEVGGLYVASRANSRRDS 320  
Qy 286 QIROSTAYSILHOPGNKEHGTGNSLYKKSDDGIRKWKOKRCSVKNGFLTISHGTANRP 345  
Db 321 QSRQG-GYSMHQLOGNKEYGSEKKFLLKKSDDGIRKWKOKRCSVKNGFLTISHGTANRP 379  
Qy 346 PAKNLNLTCCVKTNPBEKKCFDLISHORTYHFOAEDEQECQIWMVSLQNSKEEALNNAFK 405  
Db 380 PAKNLNLTCCVKTNPBEKKCFDLISHORTYHFOAEDEQECQIWMVSLQNSKEEALNNAFK 439  
Qy 406 GDNNTGNNIVQELTKIEIISVQRMTCNDVCCGAPDPTWLTSTNLGILTCIECSGIRH 465  
Db 440 GEQSTGENSL-EDLTAKIIEVQRLPGNDICCDGSGSEPTWLTSTNLGILTCIECSGIRH 498  
Qy 466 LGVHYSQMSITLDTVLGTSSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDY 525  
Db 499 MGVIHSIQLELDKLGTSSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDY 557  
Qy 526 ITAKYIERRYARKKHADNAKLSHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPD 585  
Db 558 ITAKYIERRYARKKHADNAKLSHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPD 616  
Qy 586 ETALHLAVRSVDRSLTHIVDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLLRGRKA 645  
Db 617 ETALHLAVRSVDRSLTHIVDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLLRGRKA 676  
Qy 646 SIEIANESGETPLDIKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDMD 705  
Db 677 TVDIVNGQETALDIKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDMD 736  
Qy 706 DEKLQPSNERRDRPIISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILQNETYGA 765  
Db 737 DDKSPFKKRSRPPQSFCHSSSISPD---KLALPGFSTPRDKQL-----SVGA 784  
Qy 766 L---LSSGPPPAQPAAPSTTSAPPLPRNVGK-----VQTASSANTLWKTNSV 810  
Db 785 FTNQIFASTSTDLPTSP-TSEAPPLPRNAGKGTGPPSTLPLGTQTSSGSLTKKRP 843  
Qy 811 SVDGSRQSSSDPPAVHPPLPP-----833  
Db 844 PPPFG-HKRTLSDPPSPPLPHGPPNKGAI PMGNDVGPSSSKTANKFGLSQOASTSSAKT 902  
Qy 834 -----LRVTSTN-----881  
Db 903 ALGRPVLPKLPQKVALRKTTSHLSLDRTNIPPEFQKSSQLTELQKPKPLGELPKPKV 962  
Qy 851 --AKTPSVMSEALSQPSK-----PAPPGISQIRPPP-----LPP-----881  
Db 963 ELAPKPVGELPPKPGELPPKPGQLGDLPPKPLQSLDLPKPKQMDLPPKPLQGLLAKSQA 1022  
Qy 882 -----OPPSRLPQKPAFGTKSTPLTNKGQPRGVPDLSATEALGPLSNAMVLQPPAP 934  
Db 1023 GDVSAKVQPPSEVQTR-----SHTGDLSPNVQSRDAIQKQASDESDNLTFTLP-ETVPP 1075  
Qy 935 MPRKSQATKLKPKRVKALYNCVADNPBELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRK 994  
Db 1076 LPRKINTGKVRKVTIYDQADNDDELTFIEGVEIIVTGEEDQEWIIGHIDGDPGRK 1135  
Qy 995 AFPVSFVHFAD 1006  
Db 1136 VFPVSFVHILSD 1147



## RESULT 3

A96634  
probable GCN4-complementing protein F23C21.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A96634  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A96141; MUID:21016719; PMID:11130712  
A:Accession: A96634  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-776 <STO>  
A:Cross-references: UNIPROT:Q9C6C3; GB:AE005173; NID:g10092284; PIDN:AAG12697.1; GSPDB:C  
C:Genetics:  
A:Gene: F23C21.2  
A:Map position: 1

Query Match 9.2%; Score: 485.5; DB 2; Length 776;  
Best Local Similarity 24.4%; Pred. No. 2.8e-16;  
Matches 185; Conservative 131; Mismatches 284; Indels 157; Gaps 25;  
Qy 51 LYK-MKSVKAIN--SSGLAHVENEEOYTOALEKFGNCVCRRDDPDL----GSAFLKFSV 103  
Db 34 LYGVKFKPMGALGEASTGVS-----APADSLSEEGAG--HDDPVSIVSGPVSISKIN 84  
Qy 104 FTKELTALFKNLQIONNNIISFPLDLSLKGDLKGVKGLKPPDKAWKDYETKTKIEKE 163  
Db 85 TLRELSYKFEKLRQSVHEVHLERLTNFTMTVQLQEAQ--ESRRFRDRAVHSYDQAREKPVSL 143  
Qy 164 KKEHAKLHGMTRTETSGAIEAEEMEKERRFQLOMCEVLLKVKNEIKKGVLDLQNLKY 223  
Db 144 KKN-----TRGDIV--AELEDLNSKSAFEKSRFNLSNLTAEAKKYEFLESISAI 195  
Qy 224 FQAQCNFFQDGLKAVESKPSIETL-----STDLTHTIKQAQDEERRQLIQLRDILKSL 277  
Db 196 MDSHPKYFKGLVDLSQLEPIYHQLVYAQSQKEQSKIEQDFPAQRIQEFRTQSLDSQ 255  
Qy 278 QVEQKEDSQIRQSTAYSILHQGNKEHGT-----BRNSLYKKSQDGIRKVMQKRKCS 329  
Db 256 ASAKADPSDVGGNHVYRA--IPKKNVANSVSTADKEVTKQGYLLKRSASLRADWKRRPFV 314  
Qy 330 VKN-CFLTISHGTANRPPAK-----LNLIT 353  
Db 315 LDNHGSLYYRYNTGNKSAKSHYVSGLEHSGVGFGRFTRHNRASQSGSLDCNNMIDLT 374  
Qy 354 CQVKNPEE---KKCFDLISHDRTHFOAEDEQECQIWMVSLQNSKEALNNAFKGDD-- 408  
Db 375 SLIKLDAEDTLRLCFRIISQKTYTLQAENGADRMVKNITAITRLNSHFLQOSP 434  
Qy 409 -----NTGENNIVQELT-----KEIISVQRTMGNDVCCDCAAPPTWLS 448  
Db 435 RYLDKNTSSGPATENLTNQKEDYNQRLNVDGDDVLTILRETPGNNTCAECNAPPDWAS 494  
Qy 449 TNLGTLTCEGSHRELGVHYSFMSQSLTLD--VLGTSSELLAKNIAGNAGNEIMECCLP 506  
Db 495 LNLGVMCEGSHRELGVHYSKRSUTLDVKVWEPTILDLFRLNGYCNVSWVEELH 554  
Qy 507 AEDS-----VKPNPGSDMNARKDYITAKYIERRYARKHAD--NAAKLHSLCE 552  
Db 555 HLDDSEKSGTDTLASVKPSSEDFWTKYKYNKYLEKALVVKDREANSTASSRIWE 614  
Qy 553 AVKT---RDIFGLL-QAYADGV-----DLTEKIPLANGH--EPDE----- 586

Db 615 AVQSRNIRDIYRLIVKADANIINTKFDITDLDVYHHHVDPADDEVKKRHDNACQRIKN 674  
Qy 587 -----TALHLAVRSVDRTSLSHIVDFLVQNSGLNDKQTKRGSTALHYCCLTDNAE 635  
Db 675 SNEARNCLQCGSLLHVAQSGDPI--LLELLQFGADINMRDYGRTPLHHCIAAGNNA 731  
Qy 636 CLKULLRKGAIEIETANESGETPDLQIAKRLKHEHECEL 672  
Db 732 FAKVLLRRGARPSIEDGGGLSVLERAMEMGAITDEEL 768  
RESULT 4  
T26508  
hypothetical protein Y17G7B.15 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T26508  
R:Smye, R.  
submitted to the EMBL Data Library, June 1998  
A:Reference number: Z20225  
A:Accession: T26508  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-752 <WIL>  
A:Cross-references: UNIPROT:Q9XXH9; EMBL:AL023828; PIDN:CAA19462.1; GSPDB:GN00020; CESP  
A:Experimental source: clone Y17G7B  
C:Genetics:  
A:Gene: CESP:Y17G7B.15  
A:Map position: 2  
A:Introns: 177/3; 221/3; 345/1; 539/1; 654/2; 691/2; 740/1

Query Match 9.1%; Score 479.5; DB 2; Length 752;  
Best Local Similarity 25.7%; Pred. No. 5.3e-16;  
Matches 178; Conservative 92; Mismatches 206; Indels 217; Gaps 26;  
Qy 144 KPPKAWKDYETKTKIEKEKHAHGMTRTETSGAE-----IAEMEKERRFPFQIQ 197  
Db 40 EPFSRSTKSGIRTSKV-----ITRTTISAEATKLKMDLLEQORREKR-VKLE 86  
Qy 198 M-----CEYLLKVKNEIKK-KGVLDLQNLKYFQAQCNFFQDGLKAVESKPSIETL-- 248  
Db 87 IEKSRIRAEINQRMQEIISSEPTKRGSKSLMR-----RAAKTKGRISLLI 133  
Qy 249 ---STDLTHTI-KQAQDEERRQLIQLRDILKSLAQVEQKEDSQIRQSTAYS-----LHQPQ 299  
Db 134 RDSVDSGSIITDKALKOKKRGK-----SLEWEVKDQGRFAESFSDSLMWFQHP 184  
Qy 300 QNK-EHGTFRNGSLYKKSQDGIRKVMQKKSQNGKGLTISHGTANRPPAKL---NLLTQ 355  
Db 185 GMPTEPDVMMEGYLYKRSNNAFKTNRRWFQIKDKQLLYSHRSTDLPEATIMEENLRICL 244  
Qy 356 VK---TNPEKKCFDLISHDRTHFOAEDEQECQIWMVSLQ-----NSK 396  
Db 245 VRPAPSNIDRTGCFELVPTTRIHLQADSESLCDMMRALQRTILALHEGSDVQVASTSP 304  
Qy 397 EEALNNAFKGDDNTGENNI-----VQELTK-----E 422  
Db 305 RNKTTSSSGVTLTSANNAISPLSNAMDVKGRSVSDPASTYTSANTSSISTAAGSSSTT 364  
Qy 423 IISVQRTMGNDVCCDCAAPPTWLSNLGTLTCEGSHRELGVHYSFMSQSLTLDVLG 482  
Db 365 AFEQVRVPNGEVNACDCGSPAPKWSINLGVVLTECSGAHRSIGVQTSKVRSLCMTSID 424  
Qy 483 TS--ELLAKNIGNAGNEIMECCLPADS VKP---NPGSDMNARKDYITAKYIERRYA- 536  
Db 425 NELRDVLLA--LGNRQVNEIFLAHLPADSIVPPQINEKSARPAAREAWIKAKYVRRFAV 482  
Qy 537 -----RKQH-----AD-- 542  
Db 483 AEDTRASSATNRQEHKHKHTSIGNSSNGVNRSSSYADVQDAESGGLLDADPHSADLS 542  
Qy 543 ----NAAKLHSLC-----EAVKTRDIFGLQAYADGVDLT 573

Db 543 VPVTSKSLACSGDTNLDAGSSIDTKTVEWDSVKEACECGDLLALMTAYAGGFDLN 602  
Qy 574 EKIPLANGHEPDETHALHVAHSVDRSTLSHIVDFLVQNSGNDKDGKGTSTALHYCCCLTDN 633  
Db 603 ---ALHNG---TTALHTATRNQGTAA---VEFLLNGAKINMLDEKLTPLHLAAKEGH 652  
Qy 634 AECLKLLRGKASIEIANESGETPLDIAKRLKH 666  
Db 653 TLPVCQLLKRGANDLANVDSKTPLDIAWECTH 685

RESULT 5  
T48577  
hypothetical protein T31B5.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T48577  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24490  
A:Accession: T48577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-750 <BEV>  
A:Cross-references: UNIPROT:Q9LYU6; EMBL:AL163491  
A:Experimental source: cultivar Columbia; BAC clone T31B5  
C:Genetics:  
A:Map position: 5  
A:Introns: 23/2; 57/1; 80/3; 101/3; 123/3; 139/3; 171/3; 185/3; 257/3; 334/3; 363/3; 475/3  
A:Note: T31B5.120

Query Match 8.8%; Score 466; DB 2; Length 750;  
Best Local Similarity 22.8%; Pred. No. 2.4e-15;  
Matches 180; Conservative 122; Mismatches 302; Indels 186; Gaps 26;

Qy 40 IEEALDVRMVLVYKMKSVKAINSSGLAHV-ENERQYQALEKFGNGVCVRDDPDL---- 94  
Db 1 MEESAELIURSLKPYKCRKY-TEGLGEAYDGDIAFASALETFGG--HNDPISVAFG 56  
Qy 95 GSALFVSVFTKELTALPKNIQTNNNIISFPLDSLGLKDLGVKGLKPKFPDKAWKDYE 154  
Db 57 GPVMTKFTIALREIGTYKEVLRSQVEHILNDRLLQFANWDLHEVKA-----R 103  
Qy 155 TKITKIEKEKEHAKLHGKIMRTISGABIEAEMEKERRFPQLQMCYELLYKNEIKIKGV 214  
Db 104 EKFLSLRGTGSDV-----AAALEOELHTSRSMFEQARFNLTALSNVEAKKRF 152  
Qy 215 DLLQNLIKYFHAQCFFODGLKAVESLKPSITETSLDLHTTIKQAODEBERRQLIQLRDLK 274  
Db 153 BFLEAVSGTMDAHLRYFKQGYELLHQMEPYINQVLT-----YAOQSRSRNYEQOALNE 206  
Qy 275 SALQVEQKEDSIQSTAYSILHQPQCNKEHGHTERNQ-SLYK-----KSDGIRK 321  
Db 207 KMQEYKRVQDRESRWGNSGNSPGND--GIQALGRSHKWDIAVMQSAARGKSGSQL 263  
Qy 322 VWQKRKCSVKGFL-----TISHGTANRPPAK-----LNLTCQVKNPDEE---KKCFD 367  
Db 264 SGQRNSSELGSLGRWLSNNHGGVHDEKSVARHTVNLITSTIKVDADQSLRDFCFR 323  
Qy 368 LISHDRTYHFOADEQEQIWM-----SVLQNS-KEEALNNAFKGDDNTGNNIVQOE 418  
Db 324 IISPTKNTLOAESALDQMDWIEKITGVIASLLSSQVPEQRLPGSPMG---SGHRSASE 380  
Qy 419 -----LTKELIISE-----VORMTGNDVCC 437  
Db 381 SSSYESSEVDHPTTEFVCERSFLGNERPSPSPQRSIRKGEKPIDALRKVCNDKCA 440  
Qy 438 DCGAPDPWLSNLGILTCIECSGTHRELGVHYSQMOSLTLD--VLGTSSELLAKNIGNA 495  
Db 441 DCGAPEPDWASLNLGVLCIECSGTHRELGVHYSQMOSLTLD--VLGTSSELLAKNIGNA 500  
Qy 496 GFNEIMECCLPADSVKPNPG-----SDM-NARKDIYITAKYIERRYAR 537

Db 501 FANTVWEELLHRSIAIHFDGFLTVSDKSRVMTGKPSYADMISIKYIQAKYAEKLPVR 560  
Qy 538 KGHADN--AAKLHSLCAVTRD---IFGLL-----QAYADGVDLTEKIPLAN 580  
Db 561 RSRSDFPQSAQQQMDAVSGNDKKAVYRLIVNGDADVYVYDQTSSSLTSLSVILYPE 620  
Qy 581 GHEPDETHALHVAHSVDRSTLSHIVDFLVQNSGNDK 616  
Db 621 RPKREDVLLRLNELLDRTGSSSNISPEGSGSLLHCACEKADLGHVWELLQYGANVNA 680  
Qy 617 QTGKGSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKRLKHCHCEBELLTQA 676  
Db 681 SSSGQTPHCCLLRGKVTIARULLLTRGADPEANNRSGKTALDIAESNFTDPEVL---A 737  
Qy 677 LSGRFNSHVH 686  
Db 738 LLSDTNGYNH 747

RESULT 6  
D86242  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: D86242  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroo, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719; PMID:11130712  
A:Accession: D86242  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-531 <STO>  
A:Cross-references: UNIPROT:O04097; GB:AB005172; NID:g1931654; PIDN:AAB65489.1; GSPDB:GN C:Genetics:  
A:Map position: 1

Query Match 7.6%; Score 399; DB 2; Length 531;  
Best Local Similarity 24.2%; Pred. No. 2.8e-12;  
Matches 137; Conservative 85; Mismatches 193; Indels 152; Gaps 15;

Qy 212 KGVLLQNLIKYFHAQCFFODGLKAVESLKPSITETSLDLHTTIKQAODEBERRQLIQLRD 271  
Db 3 QGYDLLNQLEPFIHQITYAQO-----SKEQSKIEQDRLARRIQEFRTQS 47  
Qy 272 ILKSALQVEQKEDSQIR-----QSTAYSL-----HOPQGNKEHGT----- 306  
Db 48 ELDSQQLVANAESSGANGNRVGGNI PYKNTETSLTADKEGSHHHVSGSDHNTGVGFRFR 107  
Qy 307 ---ERNGLYKKSIGIRKWKQKRSKVNQGLTISHGTANRPPAKNLTLTQVKNPDEE- 362  
Db 108 ARHNRSGSLTSGSLGYNTI-----DURTSILKLDADM 140  
Qy 363 ---KKCFDLISHDRTYHFOADEQEQIWMVSLQNSKEBALN----- 401  
Db 141 DLRLCFRIISPKQYTTTQAEAGADMMDWVNKITVIGTLLNSHFLQSPVRYLDKDNSS 200  
Qy 402 ---NAFKGDDNTGNNIVQELTKEIISVQRMTGNDVCCGAPDPWLSNLGILTCTIE 458  
Db 201 APANAVVSGDI RHNSRFQNIQDGVSTILRGLPGNACAECAEPDWAASLNLGVLLCIQ 260  
Qy 459 CSGTHRELGVHYSQMOSLTLD--VLGTSSELLAKNIGNAGNEIMECCLPADSV----- 511  
Db 261 CSGVHRLNGLVHISKVRSLSLDVKVWEPTILDPRNLGNVYCNLSWEGLHLLDDDCEDGSA 320

QY 512 -----KPNPGDMNARKDYITAKYIERARYARKKHAD-NAAKHLHSLCEAVKT---RDIF 560  
DB 321 LSHASVSKPCBEDSPFVKEKYLILGKYLEKALVYKIDSEANLSAASGRINEAVQSRNIREY 380  
QY 561 GLQAYAD-----GVDLTKIPLANGHEPDE----- 586  
DB 381 RLIVITGDVNIINTKFDIDITDIDAYHIDAABK-AVKKRDHTVCQRIKESNEPRSCLOG 439  
QY 587 -TALHLAVRSVDRTHLHVDFVQNSGNDKQTKGSGTALHYCCITDNNACLKLLLRGA 645  
DB 440 CSLHLVACHIGDSV---LLELLQFGADLNIRDYHGRTPLRHCISGNHKKFALLRRGA 496  
QY 646 STEIANESGETPLDIATKRLKHEHCEBL 672  
DB 497 RPSIEDDGLSVLERAMGAITDEEL 523

## RESULT 7

T49496  
hypotheical protein B14D6.480 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49496  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49496  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1087 <SCH>  
A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.480  
A:Experimental source: BAC clone B14D6; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B14D6.480  
A:Map position: 6  
A:Introns: 67/3; 99/3; 270/2; 589/2; 616/2

Query Match 7.5%; Score 393.5; DB 2; Length 1087;  
Best Local Similarity 21.7%; Pred. No. 1.3e-11;  
Matches 148; Conservative 119; Mismatches 233; Indels 181; Gaps 24;

QY 34 RNTVAAIEALDVRMVLVYKMKSKVKAI-NSSGLAHVENEEOYTOALEKPE-----GGNCV 87  
DB 334 RATMKALEQKQTNMRM---QMKRLIKRAEQVYAAQTEANDAFLOFMEALRDVYSTNANAV 390  
QY 88 CRDDPDLGSAFLKFSVFTKELTALFNKLIQNMNIIISPLDLSLLKGLDVGKGLKCPD 147  
DB 391 ---QPAIEHYFKIA---WEILRFRONTQNLQKIVIDPMNKLYQLDIIKQAEAK-KRDFE 443  
QY 148 KAWKDYETKITKIEKEKEHAKLHGIMIRTEISGABIAEBEMEKERRFFQLOMCCEYLLKVN 207  
DB 444 BESKDFYAVSVYLQGR-----HDSVKAQSDS---KYQTKRKNFELKRPDYSSFMQD 493  
QY 208 IK-IKKGVDDLQNLKIKYHQAQCNFPQDGLKAVESLKPSIETLSTDLHTIKQAQDERROL 266  
DB 494 LSGGRKEQEIILSHLTAKYADAQAKAFLTTAKYVEGLLPQLEALSTEVEQADKTYQYQRRR 553  
QY 267 IOLRDLKLSALQVEKEDSIQSTAYSILHQPQGNK-----BHGTNRNGLYK--- 314  
DB 554 BEKRRLE-----KGNLAYAPEPPTSSGSPSTGQNGTAKGGLLWALNRP 600  
QY 315 -----KSDGIRKVMQKRCVKNGLFTLISHGTANR-----PPAKNLNLT 353  
DB 601 GGHVDPRLANKQGWKFW-----IIVLDQKLSYNSWNKQRLDLHMDPIDLRAS 649  
QY 354 CQVKTNPBEKKFDLISHDRTHVFOADEQECQIWMVSLQNSKEAL----- 400  
DB 650 VREARNAERRFCFEVITPQKRYQATSEEDWNWIMAINNALQSAVEGRGLKMKPPPS 709  
QY 401 -----NNAFKGDDNTGENNIVQELTK----- 421  
DB 710 SIGHSDFDRRDRIASVLTKGSSINHHHHFHGHNSQNLSPISRRRTTVGARPSTARAPSF 769

QY 422 -----EIISEVORWITGNVCCCGAPDPT-WLSTNLGILTCIECSGIRHRELGVHYS 472  
DB 770 DESPRLQLQMLRENDQ--GNSWCADCGSSNKVWVSLNLAIIVCISGIRHSLGTHISK 827  
QY 473 MOSLTLDVLGTS---ELLAKNIGNAGFNEIMECCLPADSDSVKPNPGSDMMARKDYITA 528  
DB 828 VRSULTDITTSFTPDIIELML--VGNRVSNVWYEARLPDQKLVFQASREQRLR--FITA 883  
QY 529 KYIERY-----ARKKHADNAAKHLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLAN 580  
DB 884 KYVDRAFVEPISSTLSRYPTAE-----ETLLAAIKKEBIQQVIYALA-----LKAN 929  
QY 581 CHEPDET-----ALHLAVRSVD 597  
DB 930 PNITDKSRGTHAVYLALAAAD.950

## RESULT 8

T39731  
cbx2 protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Jul-2004  
C:Accession: T39731  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21874  
A:Accession: T39731  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-870 <WO>  
A:Cross-references: UNIPROT:Q9UUE2; EMBL:AL109846; PIDN:CAB52806.1; GSPDB:GN00067; SPDB:  
A:Experimental source: strain 972h; cosmid c17G9  
C:Genetics:  
A:Gene: SPDB:SPBC17G9.08c  
A:Map position: 2  
A:Introns: 13/1; 57/3; 118/3; 166/3; 709/1

Query Match 6.4%; Score 335; DB 2; Length 870;  
Best Local Similarity 22.2%; Pred. No. 7e-09;  
Matches 147; Conservative 111; Mismatches 254; Indels 150; Gaps 27;

QY 34 RNTVAAIEALDVRMVLVYKMKSKVKAI-NSSGLAHVENEEOYTOALEKFGNCVCRDDP 92  
DB 256 RATILNYERTTHDMRMVLKCLIKRIEHAHSHGLLYMSYKE-LMSAFER-----VATINP 309  
QY 93 DLGSAPL-----KFSVFTKELTALFNKLIQNMNIIISPLDLSLLKGLDVGKGLK 143  
DB 310 PAKPFLDHYAAQAQSFDSFNIDRARLLRFLIE-----PLRKIVDTDINVV-STKK 361  
QY 144 KPPDKAWDYETKITK-IEKEKEHAKLHGIMIRTEISGABIAEBEMEKERRF-----FOL 196  
DB 362 KDFEETSRYDTVTSLSRYLSKSEK-----ETSSDKTKESKFAAKKRDPEL 405  
QY 197 QMCVEYLLKVNIEK-IKKGVDDLQNLKIKYHQAQCNFPQDGLKAVESLKPSIETLSTDLHT- 254  
DB 406 SRFDYYSYMODINGRKGQEVLSVLTSPAANDYNIHSLTDIDALRPSIIQLQ-DIVTE 464  
QY 255 -----IKQAODEERRQLIQLRDLKLSALQVEKEDSOIROSATYSLHQPQGNKEHGT 308  
DB 465 ANKEPQLURAEERREYYIE-----TSSRELEDK-DAEI-AAQAYKAVDQDT-----SAK 513  
QY 309 NGSL--YKSKDGIRKV-----WQKRCVQNGFLTISHG-----TANRPPAKL 349  
DB 514 QGLLAFPSKSTDLQVVGKSGWHKY-----WVVDLHGKICEYANWQSLHTEPIDL 566  
QY 350 NLLTCQVKTNPBEKKCFDLISHDRTHVFOADEQECQIWMVSLQNSKEALNNAFKGDD- 408  
DB 567 LMATVRPAQSVSRKFCFEVITPQTKRYQATSKAEMHSHWIEAIOYSISESIQVKGKGTSM 626  
QY 409 NTGENNIVQELTKETISEVORM-----TGNDVC 436  
DB 627 NSEETSVKHGPTSTIGKALQVRASVTSPSRHNSDSKKQTKSPSLVLTLEKMHSSDQSC 686



QY 612 GNLDKQTKGSTALHYCCLTDAECLKLLRKGASIEIANESGETPLDIAKRLK-HEHCE 670  
Db 822 -----TKDGRVTLHLAASIGSVELAQLLIWHNADAQILDNNGRSCLFYARSGNPREVFD 875  
QY 671 ELLTQALSGRFSNVHVVEYEWRLLEDDESDDDMDKLOPSENRRDRPISFYOLGSNQ 730  
Db 876 MLVTAGLSPDYG-----LPQINDFSQMP-----FFANGSTS 908  
QY 731 LQSNVAVSLARDAANLAKEKQAFMPSIILQNETYGA-LLSGSPPPPAQAPASTTTSAPPLPP 789  
Db 909 NRDYSTS-----GDEMYGSRIRISMGPPQVPARRYLPOQPDELDE 947  
QY 790 RNW 792  
Db 948 TSV 950

## RESULT 11

T26737  
hypothetical protein Y39A1A.15a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T26737  
R:Wall, M.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z20257  
A:Accession: T26737  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-975 <WIL>  
A:Cross-references: UNIPROT:Q9XK14; EMBL:AL031633; PIDN:CAA21026.1; GSPDB:GN00021; CESP:  
A:Experimental source: clone Y39A1A  
C:Genetics:  
A:Gene: CESP:Y39A1A.15a  
A:Map position: 3  
A:Introns: 105/3; 135/1; 211/3; 257/1; 454/3; 496/2; 870/3; 924/3

Query Match 6.1%; Score 319.5; DB 2; Length 975;  
Best Local Similarity 21.6%; Pred. No. 4.6e-08;  
Matches 143; Conservative 83; Mismatches 226; Indels 211; Gaps 21;  
QY 232 QDLKAVESLKPISITLSTDLHTIKQAQDEERQILQRLDILKSALQVEQEDSQIROST 291  
Db 421 QSGANQVSGASTSHLTPSSSTPNT--QRKNRISNIFROKD-----HQBEKSKWIESL 470  
QY 292 AYSLHQPOGNEHGHGTERNGSLYKKS--DGIRKVMOKR-KCSVNGFLTI----- 337  
Db 471 NLGIRAPIKQ-----GNLYKSSKSALENKRWKKYVCLYSDGRLTYHTNLKEYMDKT 524  
QY 338 -----SHGTAN----- 343  
Db 525 AHGKEMDLKATIRITGRHLPLSHRVASSVDPMNGGGGGTPTLKSVEPRSDVGANS 584  
QY 344 -----RPPA-----KLMLTQCQVKNPEEKCFDLISHD 372  
Db 585 DGTSGGSDDAIKENORHFPPAMPQTVAKKKRESRKIGTNSKHNDDECEFEVINNC 644  
QY 373 -RTHFQAEQECQIWMVSLONSKEEALNNAFKGDDNTGNNIVQELTKHIISSVQRM 431  
Db 645 LMRWEFCAGSLEERDEWIAIGGEIEKSLG---KEVANAKTNRA-VANRPDIAALRSIP 700  
QY 432 GNDVCCDGPDPPTLWNLGILTCIEGSHIRGLGVHVSPOSITLVLGTSSELLAKN 491  
Db 701 GNGRCADCNFSSEWASNLGIIICIEGSHIRGLGVHVSPOSITLVLGTSSELLAKN 760  
QY 492 IGNAGFIMEECLPAEDSVKPNPGSDMNARKDYITAKYIERRYARKKHADNAAKHLSLC 551  
Db 761 IGNDKANENWERGL-LNGERKPTPSSREERKFRIDRVQKAFI-KPIASGPTVSLI 818  
QY 552 EAVKTRDIFGLLOAVDGLTEKIPLANGHEPDBDTALHLAVRSVDRTHLHIVDFLVONS 611  
Db 819 SAVLARDVMSL-----NVLLANG-----MSVEEINTT----- 845

QY 612 GNLDKQTKGSTALHYCCLTDAECLKLLRKGASIEIANESGETPLDIAKRLK-HEHCE 670  
Db 846 -----TKDGRVTLHLAASIGSVELAQLLIWHNADAQILDNNGRSCLFYARSGNPREVFD 899  
QY 671 ELLTQALSGRFSNVHVVEYEWRLLEDDESDDDMDKLOPSENRRDRPISFYOLGSNQ 730  
Db 900 MLVTAGLSPDYG-----LPQINDFSQMP-----FFANGSTS 932  
QY 731 LQSNVAVSLARDAANLAKEKQAFMPSIILQNETYGA-LLSGSPPPPAQAPASTTTSAPPLPP 789  
Db 933 NRDYSTS-----GDEMYGSRIRISMGPPQVPARRYLPOQPDELDE 971  
QY 790 RNW 792  
Db 972 TSV 974

## RESULT 12

T38398  
hypothetical GTPase activating protein for Arf protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T38398  
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z21791  
A:Accession: T38398  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-923 <MCL>  
A:Cross-references: UNIPROT:Q10165; EMBL:Z69240; PIDN:CAA93233.1; GSPDB:GN00066; SPDB:SF:  
A:Experimental source: strain 972h-; cosmid 26A3  
C:Genetics:  
A:Gene: SPDB:SPAC36A3.10  
A:Map position: 1  
A:Introns: 6/1; 47/3

Query Match 5.9%; Score 309.5; DB 2; Length 923;  
Best Local Similarity 20.5%; Pred. No. 1.3e-07;  
Matches 154; Conservative 116; Mismatches 243; Indels 237; Gaps 32;  
QY 13 ETHEDYKAPTASSFTTRTAQ-----CR---NTVAAIEBALDVRMVLYKMKSVKAI 61  
Db 226 ENYED--GPSFRSMLHELEQKSSLMKYCKIMKRIVQLSDAYDASQVAVMKLSETLSEA 283  
QY 62 NSSGLAHVENEEOYTOALEKEGNCVCDDPDLGSAFLKFSVFTKELTALFNLIONWN- 120  
Db 284 SNSTSMND-----ILLDSYLTAKMD-IHATFQKLN 315  
QY 121 ---NIISPLDSLKGLKGVKPKFKAWKDYETKITKIEKEKEHAKLHGMIRTE 177  
Db 316 DLINLLYEPFHNIVSSFIKPI-DDRRLEFDEQSKSFYGSLSRYLSAKDK----- 364  
QY 178 ISGAEIAEEMKERRFFQOLQCEYLLKVNKEIKIKGVDDLQNL-IKYPHAQCNFFQDGLK 236  
Db 365 -KGGD-SKFFQKXT-PALORYDYVCFMQDLHDGSIINDINGIFLOYPFHROYDHIALFSN 421  
QY 237 AVESLKSPIETLSTDLHTIK-----QAQDEERQLIQLRDILKSALQVEQEDSQIRO 289  
Db 422 LMNSVLPNLQNLKLEKTKWSTTRDRKGRMHRSQVIQTSGRPKS---MAPSPSP1-- 476  
QY 290 STAYSLHQPO----- 299  
Db 477 SPSPFLHEIQSPMPNRRMAASADDISQTSNTFTTEIKGKICISNGGASDPDKIFKEGLLVF 536  
QY 300 GNKEHGTE-----RNGSLYKKSQGIKRVKQKRCVKNGLTIS--HGT 341  
Db 537 GATELGTDLAWSKAAWHKHVIIVENGSLWEVAN-----W---KDSVKSNVSSI SLKHAS 588  
QY 342 ANRPAPKLNLLTCQKNTNPEKK--CFDLISHDRYTHFOAEDEQECQIWM----- 389  
Db 589 AD-----KVRKQGRRCFEVYVTPKJLKYQATSAEEMDSWIEACEAAKISS 635

```

Qy 390 -----SVLQNSK-----EEALNNAF--KG-----D 407
Db 636 FQLSRVATPLSASVRPSKVPLPSTSPETTPISRLSGSGIKAFPSKSGSNLQOFPRS 695
Qy 408 DNTGNNIVQ----ELTKIIESEVQMT--GNDVCCDCGA-PDPTWLSTNLGILTCIBCS 460
Db 696 DMSGTHMEQLERYHASANIFIQMLRKTDVNSVCADCGSKVDVTWCSINIPVVLICIBCS 755
Qy 461 GIHRELGVHSPMOSLTLDVLTGSELLAKIGNAGFNEIMECCLPADSDVKPNPGSDMN 520
Db 756 GIIHSLGTHISKTRSLDLSLQSKVLLCKIGNAANVRVYKGL--SNPSLKPKPEHNAQ 814
Qy 521 ARKDYITAKYIERRYARKKHADNAKLSLCEAVKTRDIFGLLOAYADGVDLTEKIPLAN 580
Db 815 VKLAFQAKKYVEHAFIDFAGVDADA---TLLEGLEQNKISKILLGLAAKPNFE-----N 866
Qy 581 GHEPDETALHLAVRSVDRSTSLHIVDFLVQN 610
Db 867 G-----VVFLKAVTRDTSKLHLELFPWN 890

RESULT 13
H59434
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence revision 03-Jun-2002 #text_change 09-Jul-2004
C:Accession: H59434; A59435; B59435; C59435
R:Billuurt, T.; Der-Sarkissian, H.; Billuurt, P.; Tissot, M.; des Portes, V.; Bruls, T.;
Eur. J. Hum. Genet. 5, 105-109, 1997
A:Title: Mapping of the X-breakpoint involved in a balanced X;12 translocation in a female
A:Reference number: H59434
A:Accession: H59434
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <B1>
A:Cross-references: UNIPROT:060890; GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
R:Billuurt, P.; Bienvenu, T.; Ronce, N.; des Portes, V.; Vinet, M.C.; Zemmi, R.; Roest C
A.; Morsine, C.; Chelly, J.
Nature 392, 923-926, 1998
A:Title: Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation
A:Reference number: A59435
A:Accession: A59435
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <B1>
A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
R:Tentler, D.; Gustavsson, P.; Leisti, J.; Schueler, M.; Chelly, J.; Timonen, E.; Amner
Eur. J. Hum. Genet. 7, 541-548, 1999
A:Title: Deletion including the oligophrenin-1 gene associated with enlarged cerebral ve
A:Reference number: B59435
A:Accession: B59435
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-802 <TEN>
A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
R:Billuurt, P.; Chelly, J.; Carrie, A.; Vinet, M.; Couvert, P.; McDonnell, N.; Zemmi, R.;
Ann. Genet. 43, 5-9, 2000
A:Title: Determination of the gene structure of human oligophrenin-1 and identification
n.
A:Reference number: C59435
A:Accession: C59435
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <B12>
A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1

```

```

Query Match 5.3%; Score 278; DB 2; Length 802;
Best Local Similarity 20.1%; Pred. No. 3.7e-06;
Matches 201; Conservative 147; Mismatches 333; Indels 318; Gaps 47;

Qy 42 EALDVRMLVYMKKSVKRAINSGLAHVENEQYTQALEKFG-----GNCVCRD 90
Db 20 ERLKCYEQLERTNKFIDKVIKDGKGNALISAMRNYSSAVKFSQTLQSFQDFDGTDLTD 79

```

```

Qy 91 DPDLGSAPLKFSEVFTKELTALFKNLIONMNNIISFPLDSLKGLKGVKGLKKPFDKAW 150
Db 80 EINTAESGFAELLINEVENERMVHNASDLLTKPLENFKEQI--GFTKRRKKKFKDQ 138
Qy 151 KDYETKITK----IEKKEHAHLGMRITETISGAIABEEMEKERRFFQLOLCEYLLKYNE 207
Db 139 ERFYSLDRHLHLSSKKKE-----SQLQEADL--QVDKERHNFPESSLDVYQIOE 187
Qy 208 IKIKGVDDLQNLKYPHAQCNFPQDGL--KAVESLKPSTETLTDLHTIKQAQDEERQ 265
Db 188 VQESKKNIVEPVLAFHL--FISNLSLTVELTDQFLPYKQQLQLSLQNTNHFSTREE 245
Qy 266 LIQLRDILKALQVEKQEDSIQROSTAYSLHQPGNKEHGTER--NGSLY---KSDGIRK 321
Db 246 MEELKKRMKEA-----POTCKLPQOPTIEGYLYTOEKWALGIS- 283
Qy 322 VWQRKGSVKNGLFTIS-----HGTANRPPAKLNLLTC--QVKTNPBEKK--CFDLISHDR 373
Db 284 -WVKYCYQYKETKTLTMTMEQKPGAKQGPLDLTKYCVRRKTESIDKRCFCFDIETNER 342
Qy 374 --TYHFOAEDEBQCIWMSVLQNSKEALNNAFKGDDNTGNNIVQELTKIIESEVQMT 431
Db 343 PGTTTLOALSSEANRLWMEAM--DGKEPIYHSPITKQOEMELNEVGFVKFCINIE--- 398
Qy 432 GNDVCCDCGAPDPTWLSTNLGILTCIBCSGTHRELGVHYSMPQSLTLDVLGTSELLAKN 491
Db 399 -----TKGIKT-----EGLYRTVGSNIQVQKLL----- 421
Qy 492 IGNAGFNEIMECCLPADSDVKPNPGSDMNARKDYITAKYIERRYARKKHADNAKLSL 551
Db 422 --NAFFDP--KC-----PGD-----VDFHNSD 439
Qy 552 EAVKTRDIFGLLOAYADGVDLTEKIPLANGHEPDET--ALHLAVRSVDRSTSLHIVDFLVQN 610
Db 440 WDIKT-----ITSSLFYLRNLSEPYMTYRLHKLKELSAK----- 474
Qy 611 SGNLDKQTKGSTALH---YCCLTDNAECLKLLRGKASIEIANESGETPLDIAKLKHE 667
Db 475 SDNLDYRLG-----AIHSLVYKLPKKNRMLLELLRHLVNV-----CE 512
Qy 668 HCEE--LLTQALSGRFSNGSHVHVEYEWRLHLEDLDESDDMDKQLQFSENREDRPISFYQL 726
Db 513 HSKENLMT-----SNMGVIFGPTLMRAQEDTVAAMNLIKFO----- 549
Qy 727 GSNOLQNAVSLAARDANLAKEKORAFWPSILQNETYALLSGSPFPAPQAPSTTSAPP 786
Db 550 -----NIVVE-----ILIEHFGKIYLG--PPEESA-----APP 575
Qy 787 LPPRNV---GKVOTASS---ANTLWKTNSVSDGGSRRQ-----SSSDP--PAVH 828
Db 576 VPPRVTARRHKPITISKRLRLERTVFTYTSLLDSEDEIQHTPNGITTSIEIEPKPPQH 635
Qy 829 PPLPLRVSTNPLTPPPPVAKTPTPSVMEALSQPSKPAP-----PGISQIRP 876
Db 636 PKLPIQRSGEDDPGRKSPSRPI-----LDGKLEPCPEVDVGVKLVSRLDQDGTKITP 686
Qy 877 P-----PLPPQPSKLPO---KKPAPGTDKSTPLTNKGQPRGPDLSATEALGPLSNAMVL 929
Db 687 KATNGPMPSGSGTTPSGFHIKRPA-----RPLAH--HKEGDAD--SFSKVRPPGKEKPTII 738
Qy 930 QPPA--PMPKRSQATKLKPKRVKALYNCVADNPDELTPS 966
Db 739 RPPVRPPDPCCRAATPOKPEKP---DIVAGNAGEITSS 774

```

```

RESULT 14
C59431
centaurin, delta 2 [imported] - human
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence revision 03-Jun-2002 #text_change 09-Jul-2004
C:Accession: C59431; D59431
R:Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Miyajima, N.; Tanaka, A.; Kotani, H.

```

DNA Res. 5, 277-286, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. XI. The complete coding sequences of 100 human genes, derived from 50 cDNA libraries and 50 genomic DNA libraries.

A:Reference number: C59431; MUID:99087487; PMID:9872452

A:Accession: C59431

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1210 <NAG>

A:Cross-references: UNIPROT:Q96L71; GB:NP\_056057; PID:g16975484; PIDN:NP\_056057.1

R;Mura, K.; Jacques, K.M.; Stauffer, S.; Kubosaki, A.; Zhu, K.; Hirsch, D.S.; Resau, J.J. Mol. Cell 9, 109-119, 2002

A:Title: ARAP1, A Point of Convergence for Arf and Rho Signaling.

A:Reference number: D59431

A:Accession: D59431

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1210 <NAG>

A:Cross-references: GB:NP\_056057; PID:g16975484; PIDN:NP\_056057.1

Query Match 5.1%; Score 271; DB 2; Length 1210;  
Best Local Similarity 24.1%; Pred. No. 1.4e-05;  
Matches 117; Conservative 74; Mismatches 178; Indels 116; Gaps 20;

Qy 211 KKGVDLLQNLIKYFHAQCNFPQDGLKAVESLK-----PSIETLSLD-LHTTIKQAOOE 262  
Db 107 KRWRLLDTLHRYFDSNKDAYSKRFISVACISHVAAGDQKFVETNNRTFAFRASDVE 166

Qy 263 RRQLIQLRDLKSAIQEOQED-SQIRQSTAYSILHQPQGNKEHGTEGSL-----YKXSD 317  
Db 167 RKEWQ-----ALQQAARARARLRSAYLLGVP-GSEQ--PDRAGSLRLGPFKNKL 216

Qy 318 GIRKVGQKRCVKVN-----GFLTISHGTANRPPAKLNLLTCQVKTNPPEKKCFD 367  
Db 217 YVAVVGDKVL-QKNLEHYHLGIGITFDMSVGNVKE-----VDRRSFD 259

Qy 368 LISHDRYTHQAEDQECQIWMVSLQNSKEALNNAFKGDDNTGNNIVOELTKELISEV 427  
Db 260 LTPYRIFSFADSELEKEQWLEMAQIAEALS-----TSEVAERI 301

Qy 428 QRMTGNDVCCDGPADPTWLSTNLGILTCESGIHRELGVHYSMPQSLTDLVLGTSSELL 487  
Db 302 WAAAPNRFACDGAQPDWASINLCVICKRAGEHRLGAGVSKVRSKMDRKVWTEYL 361

Qy 488 --LAKNTIGNAGFNEIMECCLPASDSVKPNPGSDMNARKDYITAKYIERRYAR-----K 538  
Db 362 IELFLQNGAGNRFWAANVPPEALQPS--SSPSTRCHLEAKYREGKVRVYHPLFGNQ 419

Qy 539 KHADNAKHLSCAEVKTDF---GLLQAYADV-----DLTEKIPLANGHPEDETALH 590  
Db 420 EELDKA-----LCAAVTTTDLAETQALLGCGA-GINCFSGDPEAPTPLALAEAGQTLOW 473

Qy 591 LAVRSVDRTSLHIHVDLF-----VQNSGLNDKQTKGST-----ALHY 627  
Db 474 EFLNRTTEVPRLDSMKPLEKHYSVVLPTVSHSGFLYKKTASAGKLLQDRRAREEFRRW 533

Qy 628 CCLTLD 632  
Db 534 CVLGD 538

RESULT 15  
F59430  
GTPase regulator associated with focal adhesion kinase ppl25 [imported] - human  
C:Species: Homo sapiens (man)  
C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 09-Jul-2004  
C:Accession: F59430; G59430; H59430  
R;Taylor, J.M.; Macklem, M.M.; Parsons, J.T. J. Cell. Sci. 112 (Pt 2), 231-242, 1999  
A:Title: Cytoskeletal changes induced by GRAF, the GTPase regulator associated with focal adhesion kinase  
A:Reference number: F59430  
A:Accession: F59430  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-814 <TAY>

A/Cross-references: UNIPROT:Q9UNAL; GB:NP\_055886; PID:g7662208; PIDN:NP\_055886.1  
R:Borkhardt, A.; Bojseen, S.; Haas, O.A.; Fuchs, U.; Bartelheimer, D.; Loncarevic, I.P.,  
Proc. Natl. Acad. Sci. U.S.A. 97, 9168-9173, 2000  
A/Title: The human GRAF gene is fused to MLL in a unique t(5;11)(q31;q23) and both alleles  
A/Reference number: G59430  
A/Accession: G59430  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-814 <BOR>  
A/Cross-references: GB:NP\_055886; PID:g7662208; PIDN:NP\_055886.1  
R:Xia, J.H.; Tang, X.X.; Yu, K.P.; Pan, Q.; Dai, H.P.  
Submitted to GenBank, April 2002  
A/Description: Molecular cloning of human oligophrenin-1 like (OPHNIL) gene, complete C  
A/Reference number: H59430  
A/Accession: H59430  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-814 <XTA>  
A/Cross-references: GB:NP\_055886; PID:g7662208; PIDN:NP\_055886.1

	Query Match	5.11%;	Score 267.5;	DB 2;	Length 814;
	Best Local Similarity	18.2%;	Pred. No. 1.2e-05;		
	Matches 192;	Conservative 157;	Mismatches 361;	Indels 347;	Gaps 40;
Qy	34	RNTVAATEALDVRMVLVYKKKSVKAINSSGLAHVNEEEQYTOALEKFGNCV----	CR	89	
Db	19	RELTKSHEAELDKTKPIKELIKDGKSLTSALKNLSSAKRKPADSLNEFKPQCIGDAETD	78		
Qy	90	DDPDGSAFLFSVPTKELTALPKNLIQNMNIISEPLDSSLKGLGVKGLDKKPPDKA	149		
Db	79	DEMCIARSLQEFATVLRNLEDERIRMIENASVLIITPLEKFRKEQI--GAAKEAKKKYDKE	137		
Qy	150	WKDY---ETKTKIEKEKKEHAKLHGMRITETSGABEIASBEMEXERFFQLQMCBYLLKVN	206		
Db	138	TEKCYGLEKHLNLSKKKE-----SQLEADSOVDLVKQ-HPEYVSL-EYVFKVQ	186		
Qy	207	EIKTKGVLDLONLIKYPHAQCNPFODGLKAVESLKPSTETSLDHLTIKOAQDEERQL	266		
Db	187	EVQERKNPEFVEPLLAFLQGLFTFYHHGYELAKDFGDKTQLTISIQNTRNRPFGTRSEV	246		
Qy	267	IQLRDILKSALQVEKEDSQIRQSTAYSLSHQQNKKEHGTGNSLYKSDGIRKWQKR	326		
Db	247	---ESLMKKMKENPLEHKTISPYTMGYLYVQEKRFHGTG-----WVKH	287		
Qy	327	KCSVKNGLPTIS-----HGTANRPPAKNLNLTLC-QVKTNPBEKK-CFOLISHDR--TY	375		
Db	288	YCTYQRDSKQITWVPFDQKSGKGGEDEVILKSCTRKTDSTIEKRCFDFVEADVDPGVI	347		
Qy	376	HFQAEDEQECQIWMVSLQNSKEEALNNAKPGDDNTGE---NNIVQELTKBIISBVQRMWG	432		
Db	348	TMQALSEEDRLWMEAMDG--REPVVNSKDSQSEGTALDSIGFSIIRKCIHAVETRGI	405		
Qy	433	NDVCCDCGAPPTWLSTNLGILTCIECSGIHRELGVH--YSPMQSLTLDVLGTSSELLAK	490		
Db	406	NE-----QGLYRIVGVNSRVQKLLSVLMDPKTASE-----	435		
Qy	491	NIGNAGFNEIMECCLPASDSVK-----PNPGSDMNARKDYITAKYERRRYAR	537		
Db	436	-----TETDICALWEIKTIITSALTKTYLRMLPGPLMMYQOFQSFYKAALKENQESR	485		
Qy	538	KKHADNAAKHSLCEAV--KTRDIFGLLQAYADGVDLTEKIPLANGHEPD-ETALHLAV-	593		
Db	486	-----VSEIHSVLHRLPEKNRQMLQLMNLANV-----ANNHKQNLMTVANLGVV	531		
Qy	594	-----RSVDRTSLHIVDFVQNSGNDLKQTKGSTALHYCCLTDNAECLKLLRGKASI	647		
Db	532	FGPTLLRPQEBTAAIMDKFN-----IVIEILIHNE-----	565		
Qy	648	EIANESGTEPLDIARLKHCEBELLTQALSGRFNSHVHVEYEMWLLHEDLDESDDDMDE	707		
Db	566	KIFNTVPMPLNTNAQ-----LHLSRKSSDSK--	592		
Qy	708	KLOPSENRRDRPIS-FYQLGNSQLQSNVSLARDANLAKEKQRAFMPSPILQNETYCAL	766		

Db	593	--PFS--CSRPLTLFHTVQSTEQ-----EQRNSIINSLE-----	625
Qy	767	LSGSPPPAQAAPSTTSAPPLPPRNVGKQTASSANTLWKTNVSVDGGSRQ--RSSSDP	824
Db	626	-----SVSSNP-----NSILNSSSSSLQPNMNSSD-	649
Qy	825	PAVHPPLPPLRVSTNPLTPTPPPPVAKTPTSVMEALSQPSKPAPPGISQIRPPP-	878
Db	650	-----PDLAVVRPRTRENSLPPNPSPTSPLSPS-WPMFSAPSSPMPTSTSSDSSPVRSVAG	704
Qy	879	-----LPPQPPSRLPQKKP--APGTDKSTPLTNKQOP	908
Db	705	FWMFSVAAVVLSLARSSLHAVFSLVNFVCPCHPNLHLLFDRPEEAVHEDSSTPF-----	758
Qy	909	RGVVDLSATEALGPLSNAMVLQPPAPMPRKSQATKPKRVKALYNCVADNPDELTFSEG	968
Db	759	-----RKAKALYACKAEHDSLSFTAG	780
Qy	969	DVI-IVDGEEDQEWIGHIDGDPGRKGAPPVSVVHFI	1004
Db	781	TVFDNVHPSQBFGWLEGTLN---GKTGLIPENYVEFL	814

Search completed: August 4, 2005, 08:27:21  
Job time : 36 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:07:43 ; Search time 111 Seconds  
(without alignments)  
4641.006 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGA PVPVHFIA D 1006

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5253	99.6	1006	1 DDF2_HUMAN	Q43150 homo sapien
2	4727	89.6	970	2 Q6A074	Q6A074 mus musculus
3	4215.5	79.9	1029	2 Q6NRK0	Q6nrk0 xenopus lae
4	3877.5	73.5	809	2 Q66JN2	Q66in2 mus musculus
5	3867	73.3	784	1 DDF2_MOUSE	Q78ig6 mus musculus
6	3013.5	57.1	1129	1 DDF1_BOVIN	Q97902 b 130-kda p
7	3000.5	56.9	1147	1 DDF1_MOUSE	Q9qwy8 m 130-kda p
8	2465.5	46.8	956	1 DDF1_HUMAN	Q9ulh1 h 130-kda p
9	2071.5	39.3	903	2 Q8TDY4	Q8tdy4 homo sapien
10	2067.5	39.2	903	2 Q6P9F4	Q6p9f4 homo sapien
11	1499.5	28.4	1015	2 Q7QJY3	Q7qjy3 anopheles g
12	1380	26.2	935	2 Q6AWJ6	Q6awj6 drosophila
13	1075.5	20.4	208	2 Q8N2B2	Q8n2b2 homo sapien
14	878	16.7	307	2 Q6DEH1	Q6deh1 brachydanio
15	844	16.0	762	2 Q9V367	Q9v367 drosophila
16	655.5	12.4	439	2 Q8BM35	Q8bm35 m mus muscu
17	631	12.0	424	2 Q86UY1	Q86uy1 homo sapien
18	587.5	11.1	807	2 Q6ZK55	Q6zqk5 mus musculus
19	583	11.1	307	2 Q9NXH7	Q9nxh7 homo sapien
20	582	11.0	407	2 Q9NKK2	Q9nxk2 homo sapien
21	575.5	10.9	778	2 Q6IVG4	Q6ivg4 oryctolagus
22	573.5	10.9	778	1 CBB2_HUMAN	Q15057 homo sapien
23	543	10.3	858	2 Q86UT3	Q86ut3 homo sapien
24	533	10.1	759	1 CBB5_HUMAN	Q96p50 homo sapien
25	523	9.9	833	2 Q6NXL5	Q6nxl5 mus musculus
26	515	9.8	740	2 Q8K2H4	Q8k2h4 mus musculus
27	509.5	9.7	775	2 Q8SMX5	Q8smx5 arabidopsis
28	509	9.7	740	1 CBB1_HUMAN	Q15027 homo sapien
29	504.5	9.6	843	2 Q86JDI	Q86jdi dictyosteli
30	490.5	9.3	768	2 Q9F1T8	Q9fit8 arabidopsis
31	487.5	9.2	826	2 Q9XXH8	Q9xxh8 caenorhabdi

32	487.5	9.2	826	2	Q9XZQ1	Q9xzq1 caenorhabdi
33	485.5	9.2	776	2	Q9C6C3	Q9c6c3 arabidopsis
34	481	9.1	760	2	Q6H462	Q6h462 oryza sativ
35	479.5	9.1	742	2	Q9XXH9	Q9xxh9 caenorhabdi
36	475	9.0	742	2	Q9XZQ2	Q9xzq2 caenorhabdi
37	466.5	8.8	519	2	Q6PA21	Q6pa21 xenopus lae
38	466	8.8	750	2	Q9LYU6	Q9lyu6 arabidopsis
39	455.5	8.6	799	2	Q7QFM2	Q7qfm2 anopheles g
40	449.5	8.5	552	2	Q8CI47	Q8ci47 mus musculu
41	447	8.5	202	2	Q6PJ46	Q6pj46 homo sapien
42	447	8.5	753	2	Q6ZD39	Q6zd39 oryza sativ
43	445	8.4	558	2	Q9SR80	Q9sr80 drosophila
44	441	8.4	487	2	Q6GN75	Q6gnt5 xenopus lae
45	439	8.3	828	2	Q9VCQ6	Q9vcq6 drosophila

#### ALIGNMENTS

RESULT 1  
ID DDF2\_HUMAN STANDARD; PRT; 1006 AA.  
AC O43150;  
DT 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Development and differentiation-enhancing factor 2 (Pylk C-terminus associated protein) (PAP) (Paxillin-associated protein with ARFGAP activity 3) (PAG3).  
DE activity 3) (PAG3).  
GN Name=DDEF2; Synonyms=KIAA0400;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=98116655; PubMed=9455477;  
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VIII. 78 new cDNA clones from brain which code for large proteins in vitro.";  
RT DNA Res. 4:307-313(1997).  
[2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B., Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.J., Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16999-16993(2002).  
[3]  
RP FUNCTION, PHOSPHORYLATION, INTERACTION WITH ARF1; ARF5; PTK2B  
RP AND SRC, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.  
RX PubMed=10022920;

RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,  
 RA Randazzo P.A., Schllessinger J.,  
 RT "Identification of a new Fyk2 target protein with Arf-GAP activity.",  
 RL Mol. Cell. Biol. 19:2338-2350(1999).  
 [4]  
 RN FUNCTION, MUTAGENESIS OF CYS-436, SUBCELLULAR LOCATION, AND  
 RP INTERACTION WITH PXN.  
 RX PubMed:1074932;  
 RA Kondo A., Hashimoto S., Yano H., Nagayama K., Mazaki Y., Sabe H.;  
 RT "A new paxillin-binding protein, PAG3/Papalpa/KIAA0400, bearing an  
 RT ADP-ribosylation factor GTPase-activating protein activity, is  
 RT involved in paxillin recruitment to focal adhesions and cell  
 RT migration.";  
 RL Mol. Biol. Cell 11:1315-1327(2000).  
 [5]  
 RN FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN  
 RP FILAMENTS.  
 RX PubMed:11304556;  
 RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;  
 RT "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-  
 RT ribosylation factor (ARF), regulates ARF6 in Fcgamma receptor-mediated  
 RT phagocytosis of macrophages.";  
 RL J. Exp. Med. 193:955-966(2001).  
 CC -|- FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6.  
 CC Regulates the formation of post-Golgi vesicles and modulates  
 CC constitutive secretion. Modulates phagocytosis mediated by Fc  
 CC gamma receptor and ARF6. Modulates PXN recruitment to focal  
 CC contacts and cell migration.  
 CC -|- SUBUNIT: Binds PXN, ARF1, ARF5, ARF6, PTK2B and SRC.  
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein  
 CC associated with the plasma membrane and with Golgi stacks.  
 CC Colocalizes with F-actin and ARF6 in phagocytic cups.  
 CC -|- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=PAPAlpha;  
 CC IsoId=043150-1; Sequence=Displayed;  
 CC Name=2; Synonyms=PAPbeta;  
 CC IsoId=043150-2; Sequence=VSP\_009722;  
 CC -|- TISSUE SPECIFICITY: Detected in heart, brain, placenta, kidney,  
 CC monocytes and pancreas.  
 CC -|- INDUCTION: Up-regulated during monocyte maturation.  
 CC -|- DOMAIN: The conserved Arg-464 in the Arf-GAP domain probably  
 CC becomes part of the active site of bound small GTPases and is  
 CC necessary for GTP hydrolysis.  
 CC -|- PTM: Phosphorylated on tyrosine residues by SRC and PTK2B.  
 CC -|- SIMILARITY: Contains 2 ANK repeats.  
 CC -|- SIMILARITY: Contains 1 Arf-GAP domain.  
 CC -|- SIMILARITY: Contains 1 PH domain.  
 CC -|- SIMILARITY: Contains 1 SH3 domain.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AB007860; BAA23696.2; ALT\_INIT.  
 DR EMBL; BC063308; AAH63308.1; -.  
 DR HSSP; Q60631; LGBQ.  
 DR Genew; HGNC:2721; DDEF2.  
 DR MIM; 603817; -.  
 DR GO; GO:0008047; F:enzyme activator activity; TAS.  
 DR ProDom; PD000066; SH3; 1.  
 DR PROSITE; PS00297; ANK REP REGION; 1.  
 DR PROSITE; PS00088; ANK\_REPEAT; 1.  
 DR PROSITE; PS00115; ARFGAP; 1.  
 DR PROSITE; PS50003; PH DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Alternative splicing; ANK repeat; Coiled coil; Golgi stack;  
 KW GTPase activation; Metal-binding; Phosphorylation; Repeat; SH3 domain;  
 Zinc.

FT	DOMAIN	305	397	PH.
FT	DOMAIN	421	543	Arf-GAP.
FT	REPEAT	584	616	ANK 1.
FT	REPEAT	620	649	ANK 2.
FT	DOMAIN	944	1006	SH3.
FT	DOMAIN	256	283	Coiled coil (Potential).
FT	DOMAIN	729	752	Coiled coil (Potential).
FT	DOMAIN	771	936	Pro-rich.
FT	VARSPPLIC	795	840	VQTASSANTLWKNSVSDGSRQRSSSDPPAVHPPLPPLR
FT				VTSTN -> D (in isoform 2).
FT				/FTId=VSP_009722.
FT	MUTAGEN	436		C->A: Loss of Arf-GAP activity.
FT	CONFLICT	86		C -> R (in Ref. 2).
FT	CONFLICT	748	748	E -> D (in Ref. 2).
SQ	SEQUENCE	1006 AA;	111650 MW;	6A213517DCD99E1B CRC64;
	Query Match		99.6%;	Score 5253; DB 1; Length 1006;
	Best Local Similarity		99.7%;	Pred. No. 1.2e-213;
	Matches 1003; Conservative		0; Mismatches	3; Indels 0; Gaps 0;
Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAOCNRTVAATIEEALDVRMVLVYKMKKSVKA	60	
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAOCNRTVAATIEEALDVRMVLVYKMKKSVKA	60	
Qy	61	INSSGLAHVNEEOYTQALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLQNMN	120	
Db	61	INSSGLAHVNEEOYTQALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLQNMN	120	
Qy	121	NIISFPDLSLLKGLKGVKGDLLKPKFPAKWDYETKITKEKEKEHAKLHGMIRTEISG	180	
Db	121	NIISFPDLSLLKGLKGVKGDLLKPKFPAKWDYETKITKEKEKEHAKLHGMIRTEISG	180	
Qy	181	AETAEEMEKEKRRFPQLOMCEYLLKWEIKIKKGVDDLLQNLKYPHAOCNFPQDGLKAVES	240	
Db	181	AETAEEMEKEKRRFPQLOMCEYLLKWEIKIKKGVDDLLQNLKYPHAOCNFPQDGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERRQLIQRLDKLSALQVEKEDSQIRQSTAYSLHPOQG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERRQLIQRLDKLSALQVEKEDSQIRQSTAYSLHPOQG	300	
Qy	301	NKEHGTERNGSLYKKSQDGIKRWQKRCVKNQGLTISHTGTANRPPAKLLNLTQCVKTNP	360	
Db	301	NKEHGTERNGSLYKKSQDGIKRWQKRCVKNQGLTISHTGTANRPPAKLLNLTQCVKTNP	360	
Qy	361	BEKKCFDLISHDRTHYHQAEDQECQIWMVQNSKEEALNNAFKGDDNTGENNIVQELT	420	
Db	361	BEKKCFDLISHDRTHYHQAEDQECQIWMVQNSKEEALNNAFKGDDNTGENNIVQELT	420	
Qy	421	KEIISVQRMVGTNDVCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMQSLTLDV	480	
Db	421	KEIISVQRMVGTNDVCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMQSLTLDV	480	
Qy	481	LGTSELLAKNIGNAGFNEIMECCLPADSYPKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Db	481	LGTSELLAKNIGNAGFNEIMECCLPADSYPKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Qy	541	ADNAAKLHSLICEAVKTRDIFGLQAYADGVDLTETKIPLANGHEPDETHALHVAHSVDRTS	600	
Db	541	ADNAAKLHSLICEAVKTRDIFGLQAYADGVDLTETKIPLANGHEPDETHALHVAHSVDRTS	600	
Qy	601	LHIVDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLLLRGKASTIEANESGETPLDI	660	
Db	601	LHIVDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLLLRGKASTIEANESGETPLDI	660	
Qy	661	AKRLKHEHCBELLTQALSGRFSNHSVHVYEWRLHEDLDESDDDMEKLOPSENRRDRP	720	
Db	661	AKRLKHEHCBELLTQALSGRFSNHSVHVYEWRLHEDLDESDDDMEKLOPSENRRDRP	720	
Qy	721	ISFYQLGSNQLQSNVSLARDAAANLAKEKQAFNPSPILQNETYCALLISGSPPPQAPAPS	780	
Db	721	ISFYQLGSNQLQSNVSLARDAAANLAKEKQAFNPSPILQNETYCALLISGSPPPQAPAPS	780	
Qy	781	TTSAPPLPRNVGVQVQTASSANTLWKNTNSVSDGSRQRSSSDPPAVHPPLPPLRVSTN	840	

```
Db 781 TTSAPLPPRNKVGQVQTASSANTLWKTNSVSDGSRQSSDDPPAVHPPLPLRVSTN 840
Qy 841 PLTPPPPPVAKTSPVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPKAPGTDKST 900
Db 841 PLTPPPPPVAKTSPVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPKAPGADKST 900
Qy 901 PLTNKGQPRGPVDLSATBALGPLSNAMVLQPPAPMPKRSQATKLPKPKVKALYNCAVNDP 960
Db 901 PLTNKGQPRGPVDLSATBALGPLSNAMVLQPPAPMPKRSQATKLPKPKVKALYNCAVNDP 960
Qy 961 DELTFSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006
Db 961 DELTFSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006

RESULT 2
Q6A074 PRELIMINARY; PRT; 970 AA.
ID Q6A074
AC Q6A074;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE MKTAA0400 protein (fragment).
GN Name=mkTAA0400;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Suga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-Fractionated Libraries.";
RL DNA Res. 11:205-218(2004).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC EMBL; AK172944; BAD3222.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001849; PH_related.
DR InterPro; IPR011036; PH.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3_2.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00405; REVINTACTING.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain.
FT NON_TER 1
SQ SEQUENCE 970 AA; 107768 MW; 11BAAL16234DFEECI CRC64;
```

Query Match

89.6%; Score 4727; DB 2; Length 970;

RESULT 3

```
Best Local Similarity 90.0%; Pred. No. 1.7e-191;
Matches 908; Conservative 18; Mismatches 29; Indels 54; Gaps 3;
Qy 1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Db 13 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 72
Qy 61 INSSGLAHVENEQVTALEKFGGNCVCRDDPDGLSAPLKFPSVFTKLTALPKNIQNNN 120
Db 73 INISGLAHVENEQVTALEKFGGNCVCRDDPDGLSAPLKFPSVFTKLTALPKNIQNNN 132
Qy 121 NIISFPLDSLKGLDKGVKGLKPPDKAWKDYETKITIKKEKKEHAKLHGMITETSG 180
Db 133 NIISFPLDSLKGLDKGVKGLKPPDKAWKDYETKITIKKEKKEHAKLHGMITETSG 192
Qy 181 AEIAEEMEKERRFFQLOMCEYLLKNEIKIKKGVLDLQNLIKYFHAQCNFFODGLKAVES 240
Db 193 AEIAEEMEKERRFFQLOMCEYLLKNEIKIKKGVLDLQNLIKYFHAQCNFFODGLKAVES 252
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKE---DSQIROSTAYS LHQ 297
Db 253 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKESRRDSQIROSTAYS LHQ 312
Qy 298 PQGKHEGCTERNGSLYKKSQGIKRWQKXKSVKNGFLTISHGTANRPPAKNLILTQV K 357
Db 313 PQGKHEGCTERNGSLYKKSQGIKRWQKXKSVKNGFLTISHGTANRPPAKNLILTQV K 372
Qy 358 TNPEKKCFDLISHDRTVHFOAEDEQECQIWNMSVLONSKKEALNNAFKGDDNTGNNI VQ 417
Db 373 TNPEKKCFDLISHDRTVHFOAEDEQECQIWNMSVLONSKKEALNNAFKGDDNTGNNI VQ 432
Qy 418 ELTKELIIESEVQRMGTGNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGHVYSMQS IUT 477
Db 433 ELTKELIIESEVQRMGTGNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGHVYSMQS IUT 492
Qy 478 LDVLGTSBELLAKNIGNAGFNEIMECCLPADSUKPNPGSDMNARKDVIYAKIYERR YAR 537
Db 493 LDVLGTSBELLAKNIGNAGFNEIMECCLPADSUKPNPGSDMIARKDVIYAKIYERR YAR 552
Qy 538 KKHADNAAKHSLCEAVKTRDIFGLQAYAGVDLTETKIPLANGHEPDETAHLAVRSV D 597
Db 553 KKHADTAAKHSLCEAVKTRDIFGLQAYAGVDLTETKIPLANGHEPDETAHLAVRSV D 612
Qy 598 RTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGET P 657
Db 613 RTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGET P 672
Qy 658 LDIAKRLKHEHCEBELLTQALSGRPNSHVHYEWRLHEDLDESDDDDDEKLOPSNRR E 717
Db 673 LDIAKRLKHEHCEBELLTQALSGRPNSHVHYEWRLHEDLDESDDDDDEKLOPSNRR E 732
Qy 718 DRPISFYQLGSGNOLQSNVSLARDAANLAKESQRAFMPSSIQNETYIGALLSGSPPPAOPA 777
Db 733 DRPVSFYQLGSGSQSNVSLARDTANUTKQKQKGFPSIIQNETYIGALLSGSPSSQSI 792
Qy 778 APSTTSAPPLPPRNKVGQVQTASSANTLWKTNSVSDGSRQSSDDPPAVHPPLPLRV T 837
Db 793 PPSTTSAPPLPPRNVGK----- 809
Qy 838 STNPLTPTPPPPVAKTSPVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPKAPGTD 897
Db 810 --DPLTTTPPPVAKTSGTLEAMNQPSSQSGTSQSKPPPLPPOPSPRLPQKPKASGTD 867
Qy 898 KSTPLTNKGQPRGPVDLSATBALGPLSNAMVLQPPAPMPKRSQATKLPKPKVKALYNCAV 957
Db 868 KPTPLTNKGQPRGP-----EASGPLSNAMALQPPAPMPKRSQATKSPKRVKALYNCAV 921
Qy 958 DNPDELTFSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006
Db 922 DNPDELTFSEGDVLIIDGEEDEQEWIGHIDGEPKRGKAPFVSFVHFAD 970
```

Q6NRK0  
ID Q6NRK0 PRELIMINARY; PRT; 1029 AA.  
AC Q6NRK0;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE MG83760 protein.  
GN Name=MG83760;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OC NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Fahey J.J., Helton E., Kettner M.M., Madan A.C., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RA Klein S., Strausberg R.;  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Contains 1 PH domain.  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
DR EMBL; BC070750; AAH70750.1; -.  
DR HSSP; P19878; 1KAU.  
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.  
DR InterPro; IPRO02110; ANK.  
DR InterPro; IPRO01164; ArfGAP.  
DR InterPro; IPRO01849; PH.  
DR InterPro; IPRO11036; PH related.  
DR InterPro; IPRO01452; SH3.  
DR Pfam; PF00023; Ank; 2.  
DR Pfam; PF01412; ArfGAP; 1.  
DR Pfam; PF00169; PH; 1.  
DR PRINTS; PR01415; ANKYRIN.  
DR PRINTS; PR00405; REVINTRACTING.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00105; ArfGAP; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 2.  
DR PROSITE; PS50297; ANK\_REPEAT; 1.

DR PROSITE; PS50115; ARFGAP; 1.  
DR PROSITE; PS50003; PH DOMAIN; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW ANK repeat; SH3 domain.  
SQ SEQUENCE 1029 AA; 114129 MW; BCSFCS7D0044A8C5 CRC64;  
Query Match  
Best Local Similarity 79.9%; Score 4215.5; DB 2; Length 1029;  
Matches 806; Conservative 85; Mismatches 113; Indels 27; Gaps 7;  
Qy 1 MPDQISVSEFVATHEHYDKAPTASSFTTRTAQCNTVAAIBEALDVRMVLVYKMKSVKA 60  
Db 1 MPDLISVSEFVCTNEDYKSTASNTNRLSQCRNTVVAIDEALVDHTILNKMKSVKA 60  
Qy 61 INSSGLAHVNEEQYTOALEKFGNGCYCRDDPDPLGSAFLKESVFTKELTALFKNLQNMN 120  
Db 61 IHTSGLTHVNEEQYTOALEKFGNGCYCRDDPDPLGTAFLKFSVFTKELTALFKNLQNMN 120  
Qy 121 NIIFPPLDSLKGLDKGVKGLKPPDKAWDKYTKTKIEKKEKHAHGMIRTEISG 180  
Db 121 NIITFPDLSLKGLDKGVKGLKPPDKAWDKYTKTKIEKKEKHAHGMIRTEISG 180  
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVNBIKIKGGVLLQNLIKYFHAQCNPFDGLKAVES 240  
Db 181 AEIAEEMEKERRFPQLOMCEYLLKVNBIKIKGGVLLQNLIKYFHAQCNPFDGLKAVET 240  
Qy 241 LKPSIETLSTDLHTIKQAQDEERQLQLDRILKSALQVEQKESQIRQSTAYSILHQPOG 300  
Db 241 LKPSIEKLSTDLHTIKQVDEERKQLSRLDKAALQVEQKESQIRQSTAYSILHQPOG 300  
Qy 301 NKEHGTERNGSLYKKSIGIRKWKVKSCVQNGFLTISHGTANPPAKNLNLTQCVKTNP 360  
Db 301 NKEHGTERDGYLYKKSIGIRKWKVKCTVNGFLTISHGTANPPAKNLNLTQCVKTNP 360  
Qy 361 EEKCFDLISHDRTYHFQADEQECQIWMVSLQNSKEEALNAPKGDNDTENNIVBELT 420  
Db 361 EEKCFDLISHDRTYHFQADEQECQIWMVSLQNSKEEALNAPKGDHSGENNIVBELT 420  
Qy 421 KEIISVQRTMGNDVCCDGGAPDPTWLTNLTGILTCIECSGIHRELGVHYPMSQSLTLDV 480  
Db 421 KEIISDIQRMPGNDVCCDGGASDPTWLTNLTGILTCIECSGIHRELGVHYSRIOSLTLDV 480  
Qy 481 LGTSELLLAKNIGNAGFNEIMECCLPADSDVKPNPGSDMNAKDYITAKYIERYARKKH 540  
Db 481 LGTSELLLAKNIGNSGFNEWEACLPADSDSLKPTPCSDMNAKDYITKYIERKYARKY 540  
Qy 541 ADNAAKLHSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHIAVSRDRTS 600  
Db 541 HDNASKLHLGLCDAVKSRDIPALVHLHAEGVDLTETAIPLANGHEQGETALHLAVSRDRTS 600  
Qy 601 LHIYDFLVQNSGNDLKQTKGSTALHVCCLTDNAECLKLLRGKASTEIANESGETPLDI 660  
Db 601 LHIYDFLVQNSGNDLKQTKGSTALHVCCLTDNAECLKLLRGKASTDIANDAGETPLDI 660  
Qy 661 AKRLKHEHEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDKEL-OPSENRRDR 719  
Db 661 AKRLKHAQCEDLILQAMNGRNSHVHVEYEWRLHEDLDESDDDVDEKQSPPIRREER 720  
Qy 720 PISFYQLSGNQLQSNVSLARDAANLAKERAFMPSILQNETYCALLSGSPPPAQAAP 779  
Db 721 PVSFYQLGNGQPSVACLAREAAASIAKDKPRNLPNIIQNETYGAIVGNSNSFGNAAGP 780  
Qy 780 STTSAPLPPRNVGK-----QTASSANTLWKNSVSVDCGSRQRSSDPPAVHPP 830  
Db 781 ATTNAPLPPRNTTIKVHSTIFGSSIHITPSTNPAAKASSIAESGSRQRSSDPPSIQPP 840  
Qy 831 LPPLRVSTNPLTPTPPVAKTPSVMEALSQSPKAPP---GISQIRPPPLPPPPSRL 887  
Db 841 VPPLRVST-ATNPPPPPPVTKTASVMEALSQHSKQSPPPPPVPLNKPVPVPVQPANRP 899  
Qy 888 PQKPAFCTDKSTPLTNKQGRGPVDLSATEAL-----GPILSNAMV-LQPPAPM 935  
Db 900 SHKRPAFGIDKPA-ATTKGPPPALVTASLPAADFPSETTSSQPGPKSLTPSQPPAPM 958

QY 936 PRKSOATKLKPKRVKALYNCAVADNPDELTFSEGDIIVDGEDEQEWIGHIDGPRKGA 995  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 959 PRKSQPKPKPKRVKAIYHCSADNPDELTFSEGEIIVDGEDEKQEWIGHIDGPNRGA 1018  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 996 FVPSFVHFIA 1006  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 1019 FVPSFVHFIAE 1029

## RESULT 4

O66JN2 PRELIMINARY; PRT; 809 AA.  
AC O66JN2;  
DT 25-OCT-2004 (TremBLrel. 28, Created)  
DT 25-OCT-2004 (TremBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)  
DE Gene model 592.  
GN Name=Gn592;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raba S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Director MGC Project;  
RL Submitted (AUG-2004) to the EMBL/GenBank/DBSJ databases.  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
DR EMBL; BC080847; AAH0847.1; -.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR00108; Neu\_cyt\_fact\_2.  
DR InterPro; IPR011036; PH-related.  
DR InterPro; IPR01452; SH3.  
DR InterPro; IPR011511; SH3\_2.  
DR Pfam; PF00023; ANK; 2.  
DR Pfam; PF00018; SH3\_1; 1.  
DR Pfam; PF07653; SH3\_2; 1.  
DR PRINTS; PR01415; ANKYRIN.  
DR PRINTS; PR00499; P67PHOX.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 1.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR ANK repeat; SH3 domain.  
KW ANK repeat; SH3 domain.  
SQ SEQUENCE 809 AA; 90021 MW; B77505164F392CCA CRC64;

Query Match 73.5%; Score 3877.5; DB 2; Length 809;  
Best Local Similarity 75.8%; Pred. No. 9.5e-28;  
Matches 763; Conservative 18; Mismatches 25; Indels 197; Gaps 3;  
QY 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKSVKA 60  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKSVKA 60  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 61 INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGSAPLKFPSVFTKELTALFKNLQNN 120  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 61 INISGLAHVENEEOYTQALEKFGNCVCRDDPDGLGSAPLKFPSVFTKELTALFKNLQNN 120  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 121 NIISFPLDSLKGLKGVKGLKPKFPAKWDYETKITKEKEKEHAKLHGMTRTETSG 180  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 121 NIISFPLDSLKGLKGVKGLKPKFPAKWDYETKITKEKEKEHAKLHGMTRTETSG 180  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 181 ABIAEMEKEKERRFFQLOMCEYLLKNEIKIKKGVDDLQNLKIKYHQAQCNFFQDGLKAVES 240  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 181 ABIAEMEKEKERRFFQLOMCEYLLKNEIKIKKGVDDLQNLKIKYHQAQCNFFQDGLKAVES 240  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQRSTAYSLSHQPG 300  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQRSTAYSLSHQPG 300  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 301 NKEHGTERNGSLYKSGDGIKRWOKRCSVKNQGFLLTISHGTANRPPAKLNLITCQVKNP 360  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 301 NKEHGTERNGSLYKSGDGIKRWOKRCSVKNQGFLLTISHGTANRPPAKLNLITCQVKNP 360  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 361 EEKCKFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEBALNAPFKDDNTGNNIVQELT 420  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 361 EEKCKFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEBALNAPFKDDNTGNNIVQELT 420  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 421 KEIIEVQRMGTNDVCCDGPDPDTLSTNLGILTCIECSGIRHRELGVHYSPMQSLTLDV 480  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 421 KEIIEVQRMGTNDVCCDGPDPDTLSTNLGILTCIECSGIRHRELGVHYSPMQSLTLDV 480  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 481 LGTSELLAKNIGNAGFNEIMECLPAEDSVKPNPGSDMNAKQYITAKYIERRYARKKH 540  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 481 LGTSELLAKNIGNAGFNEIMECLPAEDSVKPNPGSDMNAKQYITAKYIERRYARKKH 540  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 342 -----LAKNIGNAGFNEIMECLPSDDPKPNPGSDMIARKDYITAKYMERRYARKKH 394  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 342 -----LAKNIGNAGFNEIMECLPSDDPKPNPGSDMIARKDYITAKYMERRYARKKH 394  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 541 ADNAKHLCEAVKTRDIFGLQAYADGVDLTKIPLANGHEPDETALHLAVRSVDRTS 600  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 541 ADNAKHLCEAVKTRDIFGLQAYADGVDLTKIPLANGHEPDETALHLAVRSVDRTS 600  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 395 ADTAAKHLCEAVKTRDIFGLQAYADGVDLTKIPLANGHEPDETALHLAVRSVDRTS 454  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 395 ADTAAKHLCEAVKTRDIFGLQAYADGVDLTKIPLANGHEPDETALHLAVRSVDRTS 454  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 601 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 601 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 455 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 514  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 455 LHIIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 514  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 661 AKRLKHEHCEBELLTQALSGRFNSHVHVEYEWRLHEDLDESDDDMDKLPQSENRRDRP 720  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 661 AKRLKHEHCEBELLTQALSGRFNSHVHVEYEWRLHEDLDESDDDMDKLPQSENRRDRP 720  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 515 AKRLKHEHCEBELLTQALSGRFNSHVHVEYEWRLHEDLDESDDDMDKLPQSENRRDRP 574  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 515 AKRLKHEHCEBELLTQALSGRFNSHVHVEYEWRLHEDLDESDDDMDKLPQSENRRDRP 574  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 721 ISFYQLGSLQNSLARDANALAKEKQAFWPSILQNETYGNALLSGSPPAQPAAPS 780  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 721 ISFYQLGSLQNSLARDANALAKEKQAFWPSILQNETYGNALLSGSPPAQPAAPS 780  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 575 VSPYQLGSSQFQSNVSLARDTANLTKDKQGFPSILQNETYGNALLSGSPSSOSIPPS 634  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 575 VSPYQLGSSQFQSNVSLARDTANLTKDKQGFPSILQNETYGNALLSGSPSSOSIPPS 634  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 781 TTSAPPLPPRNVGKVTQASSANTLWKNTSVSDGSGRQRSSDRPPAVHPPLPLRVSTN 840  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 781 TTSAPPLPPRNVGKVTQASSANTLWKNTSVSDGSGRQRSSDRPPAVHPPLPLRVSTN 840  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 635 TTSAPPLPPRNVGK-----EASGLSNAMALQPPAMPKRSQATKSKPKRVKALYNCAVNDP 649  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 635 TTSAPPLPPRNVGK-----EASGLSNAMALQPPAMPKRSQATKSKPKRVKALYNCAVNDP 649  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 841 PLTTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPPQPPSRRLPQKPAPOTDKST 900  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 841 PLTTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPPQPPSRRLPQKPAPOTDKST 900  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 650 PLTTTPPPVAKTSGTLEAMNQPSKSPQSGTSQSKPPPLPPQPPSRRLPQKPAPOTDKST 709  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 650 PLTTTPPPVAKTSGTLEAMNQPSKSPQSGTSQSKPPPLPPQPPSRRLPQKPAPOTDKST 709  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 901 PLTNKGQPRGPDLSATEALGPLSNAMVLQPPAMPKRSQATKSKPKRVKALYNCAVNDP 960  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 901 PLTNKGQPRGPDLSATEALGPLSNAMVLQPPAMPKRSQATKSKPKRVKALYNCAVNDP 960  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 710 PLTNKGQPRG-----EASGLSNAMALQPPAMPKRSQATKSKPKRVKALYNCAVNDP 763  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 710 PLTNKGQPRG-----EASGLSNAMALQPPAMPKRSQATKSKPKRVKALYNCAVNDP 763  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 961 DELTFSEGDVIIVDGEDEQEWIGHIDGPRKGAFFVSVFHFIA 1006  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 961 DELTFSEGDVIIVDGEDEQEWIGHIDGPRKGAFFVSVFHFIA 1006  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
QY 764 DELTFSEGDVIIVDGEDEQEWIGHIDGPRKGAFFVSVFHFIA 809  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|  
Db 764 DELTFSEGDVIIVDGEDEQEWIGHIDGPRKGAFFVSVFHFIA 809  
||||| :|||||:|:|||||:|||||:|||||:|||||:|||||:|:|



```

Db 601 SSQIPSTTSAPPLPRNVGVK----- 623
Qy 833 PLRVSTNPLTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPPQPPSRLPQKPP 892
Db 624 -----DPLTTTPPPVAKTSGTLEANNVQSKSSQPGTSQSKPPPLPPQPPSRLPQKPP 676
Qy 893 APGTDKSTPLTNKQPGPRGVDLSATEALGPLSNAMVLOPPAPMPKRSQATKLPKRVKAL 952
Db 677 ASGADKPTPLTNKQPGRP-----EASGPLSNAMALOPPAPMPKRSQATKLPKRVKAL 730
Qy 953 YNCVADNPDELTFSEGVDVIIVDGEDQEWVTHGIDGDPGRGAPVPSFVHFIAID 1006
Db 731 YNCVADNPDELTFSEGVDVIIVDGEDQEWVTHGIDGEPGRGAPVPSFVHFIAID 784

RESULT 6
DDFL_BOVIN
ID _DDFL_BOVIN STANDARD; PRT; 1129 AA.
AC O97902;
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent Arf1 GTPase-
DE activating protein (PIP2-dependent Arf1 GAP) (ADP-ribosylation factor-
DE directed GTPase-activating protein 1) (Arf GTPase-activating protein
DE 1) (Development and differentiation-enhancing factor 1) (DEF-1).
DE Name=DEF1; Synonyms=ASAP1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 80-101; 303-324;
RP 550-558; 804-826; 839-854 AND 1039-1049.
RC TISSUE=Brain;
RX MEDLINE=99147067; PubMed=10022919;
RA King F.J., Hu E., Harris D.F., Sarraf P., Spiegelman B.M.,
RA Roberts T.M.;
RT "DEF-1, a novel src SH3 binding protein that promotes adipogenesis in
RT fibroblastic cell lines.";
RL Mol. Cell. Biol. 19:2330-2337(1999).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND SEQUENCE OF 61-65; 130-133; 424-427
RP AND 502-525.
RC TISSUE=Brain;
RX MEDLINE=99038209; PubMed=9819391;
RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
RA Randazzo P.A.;
RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that
RT associates with and is phosphorylated by Src.";
RL Mol. Cell. Biol. 18:7038-7051(1998)
CC -!- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent
CC GTPase-activating protein activity for Arf1 (ADP ribosylation
CC factor 1) and ARF5 and a lesser activity towards ARF6. May
CC coordinate membrane trafficking with cell growth or actin
CC cytoskeleton remodeling by binding to both SRC and PIP2 (By
CC similarity). May function as a signal transduction protein
CC involved in the differentiation of fibroblasts into adipocytes and
CC possibly other cell types.
CC -!- ENZYME REGULATION: Activity stimulated by phosphatidylinositol
CC 4,5-bisphosphate (PIP2) (By similarity).
CC -!- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-
CC associated; partially (By similarity).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: The PH domain most probably contributes to the
CC phosphoinositide-dependent regulation of ADP ribosylation factors
CC (By similarity).
CC -!- PTM: Phosphorylated on tyrosine residues by SRC (By similarity).
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.

```

```

CC -!- SIMILARITY: Contains 1 C4-type zinc finger.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk)
CC -----
CC EMBL; AF112886; AAD19965.1; -
CC HSSP; O89100; 1OEB.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; hRIP-like.
DR InterPro; IPR000108; Neu_Cyt_fact_2.
DR InterPro; IPR001849; PH_fact_2.
DR InterPro; IPR011036; PH-related.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00405; REVINTRACTNG.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS0297; ANK_REPEAT; 1.
DR PROSITE; PS0088; ANK_REPEAT; 2.
DR PROSITE; PS0115; ARFGAP; 1.
DR PROSITE; PS0003; PH_DOMAIN; 1.
DR PROSITE; PS0002; SH3; 1.
DR ANK repeat; Direct protein sequencing; GTPase activation; Membrane;
KW Repeat; SH3 domain; Zinc-finger.
FT DOMAIN 327 419 PH.
FT DOMAIN 442 565 Arf-GAP.
FT ZN_FING 457 480 C4-type.
FT REPEAT 603 635 ANK 1.
FT REPEAT 639 668 ANK 2.
FT DOMAIN 786 1059 Pro-rich.
FT DOMAIN 1067 1129 SH3.
FT CONFLICT 65 65 D -> T (in Ref. 2; AA sequence).
SQ SEQUENCE 1129 AA; 125381 MW; C1576C2EAC0ACD8 CRC64;

Query Match 57.1%; Score 3013.5; DB 1; Length 1129;
Best Local Similarity 55.3%; Pred. No. 3.9e-119;
Matches 627; Conservative 129; Mismatches 225; Indels 153; Gaps 23;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCNTVAATIEALDVRMLVKKKSVKA 60
Db 21 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCNTVAATIEALDVRMLVKKKSVKA 80
Qy 61 INSSGLAHVNEEYOQTALKEFGGNCVCRDDPDGLGSAPLKFSVFTKELTALPKNLQNMN 120
Db 81 IYNSGDHVQNEENYAQVLDFGNSFLSRDNPDLGTAFVKESTLTKESTLTKLNLQGLS 140
Qy 121 NIISFPLDLSLLKGLDKVGLDKKPPDKAMKDYETKITIKIEKEKHAHLGMIRTEISG 180
Db 141 HNVFTLDSLKGLDKVGLDKKPPDKAMKDYETKITIKIEKEKHAHLGMIRTEITG 200
Qy 181 AEIAEEMEKERRFPLOWCEYLLKGVVDLQNLKIYKHAQCNFPQDGLKAVES 240
Db 201 AEIAEEMEKERRFPLOWCEYLLKGVVDLQNLKIYKHAQCNFPQDGLKAVES 260
Qy 241 LKPSIETSLDHTIKQAQDEERQLQLRDLKLSALQVEQKE--DSQIRQSTAYSLSHQ 297
Db 261 LKQVIEKLADLYNIKQTQDEEKQLTALRDLKLSALQVEQKE--DSQIRQSTAYSLSHQ 319
Qy 298 PQGNKEHGTNRNGSLYKKSQDGIKRWQKRCVKNVGFITISHGTANRPPAKNLNLTQGVK 357

```



Db 320 LQGNKEYGSEKKVLLKSGDGRKVRQRRKCSVKNGILTI SHATSNRQPAKLNLLTCQVK 379  
Qy 358 TNPEKKCFDLISHDRTHFOABEQBCQIWMVLQNSKEBALNNAFKGDDNTGNNIVQ 417  
Db 380 PNAEDKKSFDLISHNRTHFOABEQDQVVAIVSVLTNSKEALTMAFRGEQSGESSL-E 438  
Qy 418 ELTKETISEVORMTGNDVCCDGGADPTWLTSTNIGLITCIKCSGTHRELGVHYSPMOSLT 477  
Db 439 ELTKAILEDVQLPNDVCCDGGAEPTWLTSTNIGLITCIKCSGTHREMGVHISRIOSLE 498  
Qy 478 LDVLGTSSELLAKNIGNAGFNEIMECCLPADSDVKPNPQSGDMNARKOYITAKYIERRYAR 537  
Db 499 LDKLGTSELLAKXVGNNSFNDIMEANLPS-PSPKPTSSDWTYKVEITYAKYVDHREPSR 557  
Qy 538 KGHADNAKLHSLCEAVKTRDIFGLLQAYAGVDLITEKIPIANGHEPDDETHALHLAVRSVD 597  
Db 558 KTCSSSAKLNELLEAIKSRDLALIQVYABGVLMPEPL-LEPGQELGETALHLAVRTAD 616  
Qy 598 RTSLHIVDFLVONSGNLDKQTKGSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 657  
Db 617 QTSUHLVDVFNQCNLQKQALGNALHYCSMYSKPECLKLLRSKPTVDVNVQAGETA 676  
Qy 658 LDIARLKHCECEILLTQALSGRNSHVHVEYEWRLLEDHEDDDMDKLPQSENRR 717  
Db 677 LDIARLKATQCEDLLSQAQSKGNPHVHVEYEWNLROEWDSDDDLLDKPSPKKERS 736  
Qy 718 DRPISFYOLGNQIQSNVSLARDAANLAKQKRAFMPISILQNETYGA-----LLSGSP 771  
Db 737 PRQSFCHSSSISPDQ---KLSLPGFSTPRDKQL-----SYGAFTNQIFVSTSD 784  
Qy 772 PPAOPAPSTTSAPLPPRNVGK-----VOTASSANTLWKTNSVSDVGGSRQR 819  
Db 785 SPTSPGIA-----EAPPLPRNATKGGPGPSTLPTSTQTSSSGSLSKKRSPPPPGPKR 839  
Qy 820 SSSPPPAVHPLPLPLR--VTSTNPLTPPPPPVAPTPSVMEALSO----- 862  
Db 840 TLDSPPSPLPHGPNKGVNWDVGFSS---SKTNNKFGLSQSQSGTSKALGPRV 896  
Qy 863 -----PSKPA-----PPGI-----SQI-----RPPQ-----LPQP----- 883  
Db 897 LPKLPQKVALRKTTSHHLSLDKANVPPEIFQKSSQLTELKPKPPGDLPPKPTLAPKP 956  
Qy 884 -----PSRLPK-----KAPAGTDKSTPLNKGOPR 909  
Db 957 PIGDLPKPGELPKPQGLDLPKQADLPKQVLDLPKQGLLAKPQGDASPK 1016  
Qy 910 G-----PVDLS-----ATEALGPLSNAMVLQPPAPMPKRSQATKLKPKRVKAL 952  
Db 1017 AQPLELTPKSHPADLSNVPKQASEDNDLTPLP-ETPVLPRKINTGSKVRVKTI 1075  
Qy 953 YNCVADNPDELTFSEGDIIVDGEDQBWHTGHIDGPRKGAPFVSPVHFAD 1006  
Db 1076 YDCQADNDELTFMEGEVIVVTGEDQEWHTGHIEGQPERKGVFVSPVHILSD 1129

RESULT 7  
DDFI MOUSE STANDARD; PRT; 1147 AA.  
AC Q9QW8; Q08612; Q80T68; Q80UV6; Q99LV8; Q942B6;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-activating protein (PIR2-dependent ARF1 GAP) (ADP-ribosylation factor-directed GTPase-activating protein 1) (ARF GTPase-activating protein 1) (Development and differentiation-enhancing factor 1) (ARF-1).  
DE 1) (Differentiation-enhancing factor 1) (DEF-1).  
GN Name=def1; Synonyms=Asap1, Kaa1249, Shag1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND  
RP MUTAGENESIS OF ARG-811; PRO-910 AND PRO-913.  
RC TISSUE=Brain, and Embryo;  
RX MEDLINE=99038209; PubMed=9819391;  
RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,  
RA Randazzo P.A.;  
RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that  
RT associates with and is phosphorylated by Src.";  
RL Mol. Cell. Biol. 18:7038-7051(1998).  
RN (2)  
RN SEQUENCE OF 57-1147 FROM N.A. (ISOFORM 3).  
RN TISSUE=Brain;  
RC MEDLINE=22579291; PubMed=12693553;  
RX Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
RA Nakajima D., Negase T., Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene  
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries.";  
RL DNA Res. 10:35-48(2003).  
RN (3)  
RN SEQUENCE OF 654-1147 FROM N.A. (ISOFORM 1).  
RN MEDLINE=97271433; PubMed=9126384; DOI=10.1006/abio.1997.2040;  
RX Yamabhai M., Kay B.K.;  
RT "Examining the specificity of Src homology 3 domain -- ligand  
RT interactions with alkaline phosphatase fusion proteins.";  
RL Anal. Biochem. 247:143-151(1997).  
RN (4)  
RN SEQUENCE OF 20-1147 FROM N.A. (ISOFORM 4), AND SEQUENCE OF 722-1147  
RN FROM N.A. (ISOFORM 1).  
RC STAFIN-C57BL/6J; TISSUE=Breast tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Deng J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S.K., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN (5)  
RN TISSUE SPECIFICITY.  
RP MEDLINE=99147067; PubMed=10022919;  
RX King F.J., Hu E., Harris D.F., Sarraf P., Spiegelman B.M.,  
RA Roberts T.M.;  
RT "DEF-1, a novel src SH3 binding protein that promotes adipogenesis in  
RT fibroblastic cell lines.";  
RL Mol. Cell. Biol. 19:2330-2337(1999).  
CC -1- FUNCTION: May function as a signal transduction protein involved  
CC in the differentiation of fibroblasts into adipocytes and possibly  
CC other cell types (by similarity). Possesses phosphatidylinositol  
CC 4,5-bisphosphate-dependent GTPase-activating protein activity for  
CC ARF1 (ADP ribosylation factor 1) and ARF5 and a lesser activity  
CC towards ARF6. May coordinate membrane trafficking with cell growth  
CC or actin cytoskeleton remodeling by binding to both SRC and PIP2.  
CC -1- ENZYME REGULATION: Activity stimulated by phosphatidylinositol  
CC 4,5-bisphosphate (PIP2).  
CC -1- SUBUNIT: Homodimer. Interacts with SRC and CRK.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-  
CC associated; partially.





Db 677 TVDIVNQNETALDIKRLKATQCEDLLSQAKSGKFNPHVEYWNLRQDEMDESDDDL 736  
Qy 706 DEKLQPSNREDDPISFYQLSGNQSNVSLARDAAANLAKERQAFWPIILQNETYGA 765  
Db 737 DDKSPFKKRSPPRSQSFCHSSISPD- ---KLALPGFSTPRDKQL- : : : : :  
Qy 766 L- ---LSGSPPPAQAAPATTAPPLPPRNVGK- ---VQTASSANTLWKTNV 810  
Db 785 FTNQIFASTDLPSTP- ---TSAPPLPPRNAGKPTGPTSTPLPLGTQTSSGSTLSKKRPP 843  
Qy 811 SVDCGSRQRSSDPPAVHPPPLP- --- 833  
Db 844 PPPPG-HKRTLSDDPSPLPHGPPNKAIPWGNVDPSSSKTANKFGLSQOASTSSAKT 902  
Qy 834 - ---LRTSTN- ---PLTPTPPPPV 850  
Db 903 ALGPRVLKPLQKVALRKTKTSHLSLDRTNWIPETFKSSQLTLPQKPLGELPPKPV 962  
Qy 851 - ---AKTPSVMEALSQPSK- ---PAPGISQIRPPP- ---LPP- --- 881  
Db 963 ELAPKPVGELPPKPGELPPKQLGDLPPKQLSDLPKPKQMKOLPPKQLGDLAKSQ 1022  
Qy 882 - ---QPPRLPQKPAQGTDKSTPLTNKGQPRGVDLSATEALGPLSNMVLQPPAP 934  
Db 1023 GDVSAKVQPPSEVTR- ---SHTGDLSPNVQSRDAIQKQASEDSNDLTPTLP-ETEPV 1075  
Qy 935 MPRSQATKLPKRVKALYNCAVDNPDLTSEGDVVIIDGDEDEWIGHIDGDPGRKG 994  
Db 1076 LPRKINTGKNVRRVKTYYDQANDDELTFTEGEVVIITGDEDEWIGHIEGQPERKG 1135  
Qy 995 APFVSFVHFID 1006  
Db 1136 VPPVSFVHILSD 1147

RESULT 8  
DDF1\_HUMAN  
ID\_DDF1\_HUMAN STANDARD; PRT; 956 AA.  
AC Q9ULH1;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-  
activating protein (PIP2-dependent ARF1 GAP) (ADP-ribosylation factor-  
directed GTPase-activating protein 1) (ARF GTPase-activating protein  
1) (Development and differentiation-enhancing factor 1) (fragment).  
GN Name=DDEF1; Synonyms=ASAP1, KIAA1249;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Bone marrow;  
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Amid C., Osanger A.,  
RA Pobo G., Han M., Wiemann S.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 11-959 FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=20039619; PubMed=10574462;  
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,  
RA Chara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:337-345(1999).  
RN [3]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=99038209; PubMed=9819391;  
RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,  
RA Randazzo P.A.;  
RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that

RT associates with and is phosphorylated by Src.";  
RL Mol. Cell. Biol. 18:7038-7051(1998).  
CC -!- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent  
CC GTPase-activating protein activity for ARF1 (ADP ribosylation  
CC factor 1) and ARF5 and a lesser activity towards ARF6. May  
CC coordinate membrane trafficking with cell growth or actin  
CC cytoskeleton remodeling by binding to both SRC and PIP2. May  
CC function as a signal transduction protein involved in the  
CC differentiation of fibroblasts into adipocytes and possibly other  
CC cell types (By similarity).  
CC -!- ENZYME REGULATION: Activity stimulated by phosphatidylinositol  
CC 4,5-bisphosphate (PIP2) (By similarity).  
CC -!- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-  
CC associated; partially (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=2;  
CC IsoId=Q9ULH1-1; Sequence=Displayed;  
CC Name=1;  
CC IsoId=Q9ULH1-2; Sequence=VSP\_008365;  
CC -!- DOMAIN: The PH domain most probably contributes to the  
CC phosphoinositide-dependent regulation of ADP ribosylation factors  
CC (By similarity).  
CC -!- PTM: Phosphorylated on tyrosine residues by SRC (By similarity).  
CC -!- SIMILARITY: Contains 2 ANK repeats.  
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.  
CC -!- SIMILARITY: Contains 1 C4-type zinc finger.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/submit/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC  
CC EMBL; BX537768; CA097831.1; -;  
CC EMBL; AB033075; BAA06563.1; -;  
CC IntAct; Q9ULH1; -;  
CC MIM; 605953; -;  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR001164; hRIP\_like.  
CC InterPro; IPR000108; Neu\_Cyt\_fact\_2.  
CC InterPro; IPR001849; PH.  
CC InterPro; IPR011036; PH related.  
CC InterPro; IPR001452; SH3.  
CC Pfam; PF00023; Ank; 2.  
CC Pfam; PF01412; ArfGap; 1.  
CC Pfam; PF00169; PH; 1.  
CC Pfam; PF00018; SH3; 1.  
CC PRINTS; PR00499; P67PHOX.  
CC PRINTS; PR00405; REVINTRACTING.  
CC PRINTS; PR00452; SH3DOMAIN.  
CC ProDom; PD000066; SH3; 1.  
CC SMART; SM00248; ANK; 2.  
CC SMART; SM00105; ArfGap; 1.  
CC SMART; SM00233; PH; 1.  
CC SMART; SM00326; SH3; 1.  
CC PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE; PS50088; ANK\_REPEAT; 1.  
CC PROSITE; PS50115; ARFGAP; 1.  
CC PROSITE; PS50003; PH\_DOMAIN; 1.  
CC PROSITE; PS50002; SH3; 1.  
KW Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;  
KW SH3 domain; zinc-finger.  
KW NON\_TER 1  
FT DOMAIN 151 243 PH.  
FT DOMAIN 266 389 Arf-GAP.  
FT ZN\_FING 281 304 C4-type.

```
FT REPEAT 427 459 ANK 1.
FT REPEAT 463 492 ANK 2.
FT DOMAIN 610 820 Pro-rich.
FT VARSPLIC 894 956 SH3.
FT VARSPLIC 130 130 E -> ESRR (in isoform 1).
FT SEQUENCE 956 AA; 105967 MW; 68CB7034EC5B8E0 CRC64;

Query Match 46.8%; Score 2465.5; DB 1; Length 956;
Best Local Similarity 53.3%; Pred. No. 4.2e-96;
Matches 520; Conservative 110; Mismatches 204; Indels 141; Gaps 21;

QY 154 ETKITKIEKEKEHEKAKLHGMIRTEISGABIAEMEKERRFFOLQCEYLLKVNKIKKG 213
Db 1 ETKFKIEKEKEHAKHGMIRTEITGABIAEMEKERRFFOLQCEYLLKVNKIKKG 60
QY 214 VDLLQNLKLYFHAQCNFFQDGLKAVESLKPSITETSLDHTIKQAQDEERRLOLIDIL 273
Db 61 VDLLQNLKLYFHAQCNFFQDGLKAVESLKPSITETSLDHTIKQAQDEERRLOLIDIL 120
QY 274 KSALQVEQKEDSQRISTAYSILHQPQGNKEHGTGERSLYKSDGIRKVMQKRCVKNK 333
Db 121 KSSLQDQKEDSQSQG-GYSMHQIQGNKEYSKGYLLKSDGIRKVMQKRCVKNK 179
QY 334 FLTISHGTANRPAPKLNLLTCQVKNPREKCFDLISHDRTHFOAEDEQECQIWMVLQ 393
Db 180 ILTISHATSNRPAPKLNLLTCQVKNPREKCFDLISHDRTHFOAEDEQECQIWMVLQ 239
QY 394 NSKEALANNAFKGDDNTGNNIVOLTEKILSEVQRMGTNDVCCDGPADPTWLNGLI 453
Db 240 NSKEALANNAFKGDDNTGNNIVOLTEKILSEVQRMGTNDVCCDGPADPTWLNGLI 298
QY 454 LTCIECSGHRHGLVHYSPMSLTLVLGTSLLAKNIGNAGNEIMECCLPADSVKVP 513
Db 299 LTCIECSGHRHGLVHYSPMSLTLVLGTSLLAKNIGNAGNEIMECCLPADSVKVP 357
QY 514 NPGSDNNARKOYITAKYIERRYARKHADNAKHLCEAVKTRDI FGLQAYADGVLT 573
Db 358 TPSSDMTVRKEYITAKYIDHFRPSRKTCTSTSAKLNELLEAKSRDLALIQVYABGVLM 417
QY 574 EKIPLANGHEPDETALHLAVSVDTSLHIVDFLVNSGNDLKOTGKSTALHYCCLTDN 633
Db 418 EPL-LEPQOELGETALHLAVRTADQTSLHIVDFLVNSGNDLKOTGKSTALHYCCLTDN 476
QY 634 AECLKLLRGKASIBIANESGETPLDIAKRLKHECEBELLTOALSGRFSNHVVEYWL 693
Db 477 PECLKLLRSKPTVDIVNAGETALDIAKRLKATQCEBLLSOAKSGKNPHVVEYWL 536
QY 694 LHEDLDESDDMDKLOPSENRRDRPISFYOLGNSQLQSNVSLARDAANLAKQKRAF 753
Db 537 RQEEIDESDDMDKLOPSENRRDRPISFYOLGNSQLQSNVSLARDAANLAKQKRAF 592
QY 754 MPSILQNETYALLSG---SPPAPAPASTTSAPLPVRVKG-----VOTA 798
Db 593 ---SYGAFTNQIFVSTSDTSPTSP-TTEAPLPFRNAGKPTGPPSTPLSTQTS 643
QY 799 SSANTLWNTSVSDGSGKRSRSDPPPAVHPPLPLR----- 835
Db 644 SGSSTLSKRRPPPPPPG-HKRTLSPPSPPLPHGPNKGAVPNGNDGPGSSSKTKNKFEG 702
QY 836 -----VTSTNP-LTPTPPPVA--KT-----PSVMEALSQ-----PSKPA 867
Db 703 LSQSSSTSSAKTALGPRVLPLKPKQVALRKTDHLSLDKATIPPEIFQKSSQLAELPQKPP 762
QY 868 P-----PGIQIRPPP-----LPQPP-SRLPQK----- 890
Db 763 PGDLPPKPTLAPKFOIGDLPPKPGELPPKPOLGDLPPKPOLSLPPKMDLPPKPOL 822
QY 891 -----KPAPQTKSTP--LTNKGQPRGPVLDLSATEALGPLSNAMVLQP 931
Db 823 GDLLAKSQTGDSVPAQOPSEVTLKSHPLDLSPNVQSRDAIQKQASENSDLTPTLP-ET 881
QY 932 PAPMPKRSQATKLKPKRVKALYNVADNPDELTFSEGDIIVDGBEDQEWIGHIDGDPG 991
```

```
Db 882 FVPLPRKINTGNKVRVKTIYDQADNDDLTFTGEVVIITGBEDQEWIGHIEGQPE 941
QY 992 RKGAPPVSVFVHFAD 1006
Db 942 RKGVPVSVFVHLS 956

RESULT 9
Q8TDY4
ID Q8TDY4 PRELIMINARY; PRT; 903 AA.
AC Q8TDY4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Up-regulated in liver cancer 1.
GN Name=UPLC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Okabe H., Furukawa Y., Kato T., Hasegawa S., Nakamura Y.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; AB056749; BAB85677.1; -.
DR HSSP; O7SIG6; 1DQC.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002034; AIPM/HcIt_synth.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW ANK repeat.
SQ SEQUENCE 903 AA; 99155 MW; D5E89518AFC86656 CRC64;

Query Match 39.3%; Score 2071.5; DB 2; Length 903;
Best Local Similarity 43.4%; Pred. No. 1.6e-79;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;

QY 1 MPDQISYSEFVAETHEDYKAPT-ASSFTTRTAQCRNTVAALIEEALDVRMVLKMKKSVK 59
Db 1 MPEQFSAEFLAVTAEDLSSPAGAAFAAKMPYRGAALAREEILEGQAILQRIKKAVR 60
QY 60 AINSGLAHVNEBQYTOALEKFGNCVCRRDDPDLSAFKPSVFTKELTALFKULIQNM 119
Db 61 AIHSSGLGHVNEBQYREAVESLGNHLSQNSHLSGTGLNLAVFTREVAALFKULIQNL 120
QY 120 NNITSPFLDLSLKGDLKGVKGLKPPDKAWDYETKITTEKEKEHAKLHGMIRTEIS 179
Db 121 NNIVSPFLDLSLKGDLKGVKGLKPPDKAWDYETKITTEKEKEHAKLHGMIRTEIS 174
QY 180 GAETAEEMEKERRFFOLQCEYLLKVNKIKKGVDLQNLKLYFHAQCNFFQDGLKAVE 239
Db 175 PGEVAQDMQREIRIFQLHMCYELLKAGESQMKQGPDLFQSLIKFHAQHNFFQDGWKAQ 234
QY 240 SLKPSIETSLDHTIKQAQDEERRLOLIDILKSLAQVEQKEDSQRIST--AYSJHQ 297
```

```
Db 235 SLFPPIEKLAAASHALHQAQDELOKLTQLRDSLRGTLQLSREBHLSRKNSGCGYSIHQ 294
Qy 298 POGNKEHGTERNGSLYKESDGIKWKQKSCVKNGLFTISHGTANRPAPKLNLLTCOVK 357
Db 295 HQGNKQFTEKVGFLYKKSGLRRVWQKKGCVKGCUTISHSTINRRFPVKLLTTCQVR 354
Qy 358 TNPBEKKCFDLISHDRTHFOAEDRQEQIWMKSVLQNSKEBALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTHFOAEDHECEAWVSVLQNSKDEALSAFLGESAPGSGWS 414
Qy 410 TGENNIVOLKEIIESEVORMTNDVCCGAPDPTWLTSTNLGILTCIECGIHRELGVH 469
Db 415 AGHDGEHDLTKLLIAEVKSRPGNSQCCDCAADPTWLTSTNLGVLTCIQCSGVHRELVR 474
Qy 470 YSPQSLTDLVLTSELLAKNAGNEIMECCLPRAEDSVKPNPGSDMNAKDYITAK 529
Db 475 FSRMQSLTDLGPESELLALNMGNSTFNEVMEAPLPSHGPKPSAESDMGTRRDYINAK 534
Qy 530 YIERRYARKKIADNAAKLHSLCEAVKTRDIFGLIQAVADGVDLTEKIKPLANGHPEDETAL 589
Db 535 YVEHRFARR-----CTPEQRRLWTAICNRDILLSLEAFANGQDFGQPLPGPAQAEELVL 590
Qy 590 HLAIVRSVDRTHLHIVDFLVQNSGNDKQTKGSTALHYCCLTDNAECLKLLRKGASIEI 649
Db 591 HLAIVKVAQASLPLVDFIIQNGCHLDAKADGNTALHYALYNQDCLKLLKGRALVGT 650
Qy 650 ANESGETPLDTAKRLKHECHCELLTOALSGRFNSHVHYEWRLHEDLDESDDMDKEL 709
Db 651 VNEAGETALDIARKKHKECEELLEQAQGTFAFPLHVDYGSWVISTEPGSDSEDEEBK- 709
Qy 710 QPSENRRDRPTSFYQLGNSQLQNSVSLAADAAMLAKEKORAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPQAQHWASGRLLDISNKTYETVASL 739
Qy 769 GSPPPAPAAPSTTSAPLPRNVKQVQTASSANTLMKNTSVSDVGGSRQSSSDPPAVH 828
Db 740 GA-----ATQGESEDCPPLPVKN-----SSRTLQV-----GCARUASGDRSEVS 780
Qy 829 PPLPLRVTSINPLTPTPPPPVAKTPSVMEALSQPSKP--APPGISQ-IRPPPLPPQ-- 883
Db 781 -----SUSSEAPETPESLGSPPASSLSPL-EPGDFSQAPPNSEEGLEPPGTSRPSL 833
Qy 884 -----PSRLPOKKPAPGTDKSTPLTNKGOPRGPVDLSATEALGPLSNAMVLQPPAPMPRK 938
Db 834 TSGTTPSEM--YLPVRFSSESTRYRG-ANSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLKPKRVKALYNCVADNPDELTFSEGDVVIIVDGEEDQEWMTGHIDGDPGRKAPPV 998
Db 877 NVP-----VGIPEGDGSRTGSLPA 895
Qy 999 SFVHFIAD 1006
Db 896 SSVQLLQD 903

RESULT 10
Q6P9F4 PRELIMINARY; PRT; 903 AA.
AC Q6P9F4;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Up-regulated in liver cancer 1.
GN Name=UPLC1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
TX TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

```
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Abramo R.D., Mullaly S.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC060786; AAH60786.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002034; ALPM/Hcit_synth.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00023; Ank; 1.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00815; ALPM_HOMOCIT_SYNTH_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR ANK repeat.
KW ANK repeat.
SQ SEQUENCE 903 AA; 99126 MW; D291DF00FD82797D CRC64;

Query Match 39.2%; Score 2067.5; DB 2; Length 903;
Best Local Similarity 43.3%; Pred. No. 2.4e-79;
Matches 445; Conservative 164; Mismatches 272; Indels 147; Gaps 21;

Qy 1 MPDQISVSEFVAETHEDYKAPT-ASSFTTRTAQCRNTVAATEEALDVRMVLVYKMKGVK 59
Db 1 MPEQFSVAEFLVTAEDLSSPAGAAFAAKMPRYGAALAREEILEGDAQILRIKKA 60
Qy 60 AINSSGLAHVENEEOYTQALEKFCGNCVCRRDDPLGSAFLKFSVFTKLTALFNKLIONM 119
Db 61 AIHSSGLGHVENEEOYREAVESLGNLSHLSTGLFLNLAFTREVALFNKLIONL 120
Qy 120 NNIISFPLDLSLLKGLDKGKDPDKAMPKWDYETKITKEKKEKHAKLHGMIRTEIS 179
Db 121 NNIVSFPDLSLMLKGLRGRQDSKKOLEKAWDYEAKKWLEKE-RDRAVVTGGI----- 174
Qy 180 GAETAEEMEKEKRRFPQLQCMCEYLLKLVNEIKTKKGVDLLQNLKIFHACNPFQDGLKAVE 239
Db 175 PGEVAQMQERRIFQLHMCVEYLLKAGESQMKQGPDLQSLIKFFHAGHNFQDQWKAQ 234
Qy 240 SLKPSIETLSTDLHTIKQAQDEERRQLQLDLKLSALQVEKEDSQSQST--AYSILHQ 297
Db 235 SLFPPIEKLAAASHALHQAQDELOKLTQLRDSLRGTLQLSREBHLSRKNSGCGYSIHQ 294
Qy 298 POGNKEHGTERNGSLYKESDGIKWKQKSCVKNGLFTISHGTANRPAPKLNLLTCOVK 357
```

```

Db 295 HGNKQFGTEKGYFKYKSGDGRRAWQRKCKGVKGLTISHSTINRPPVXLTLTCQVR 354
QY 358 TNPEKKCFDLISHDRTYTHFOADEQEQIWMVSLQNSKEEALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTYTHFOADEHECEAWVSLQNSKDEALSAFLGEPGAGSGS 414
QY 410 TGNNIVVELIYKEIISVQRTMGNDVCDGAPDPTWLNGLTICIECSGIRHRELGVH 469
Db 415 AGHDGEPHDLTKLLIAEYKSRPGNSQCCDGAADPTWLNGLTICIECSGIRHRELGV 474
QY 470 YSPMOSLTDLVLTGSELLAKNIGNAGNEIMECCLPADSCKPNPMDGMARKDYITAK 529
Db 475 FSRMOSLTDLVLTGSELLAKNIGNAGNEIMECCLPADSCKPNPMDGMARKDYITAK 534
QY 530 YIERYARKKHADNAKLSHCEAVKTRDIFGLLOAYAGVDLTETKIPLANGHEPDETAL 589
Db 535 YVEHRFARR---CTPEPQLWTAICNRDLSLVEAFANGQDFGQPLFGPDAQAEELVL 590
QY 590 HLAHSVDRSTLSHIVDFLVQNSGNDKOTGKSTALHYCCLTIDNAECLKLLLRGKASIEI 649
Db 591 HLAHVANQASLPLVDFIIONGGHLDAAKADGNTALHYAALYNQPDCLKLLKGRALVGT 650
QY 650 ANESGETPLDTAKRLKHECHCELTQALSGRPNSHVYEWRLHEDLDESDDMDKEL 709
Db 651 VNEAGETALDIAKXKHKECEBELLEQAQAGTAPPLHVDYISWVISTEPGSDSEDEBK- 709
QY 710 QPSENRRDRPISFYOLGNSQLQNSAVSLARDAANLAKORAFMPSILQNETYVGLLS- 768
Db 710 -----RCLLKLPAQAHWASGRDLISNKTYETVASL 739
QY 769 GSPPPAQAAPSTTSAPPLPRNVKGVQTASANTLWKTNSVSDGGRQRSSDPPPAVH 828
Db 740 GA---ATPQGESEDCPPPLPVKN-----SSRTLQ-----GCARUASGDRSEVS 780
QY 829 PPLPLRVTSINPLTPPPVAKTPSVMEALSQSKP--APPGISQ--IRPPPLPPQ-- 883
Db 781 -----SUSSEAPETPESIGSPASSSLMSPL-EPGDPSPAPNSSEGLREPPPGTSRPSL 833
QY 884 -----PSRLPQKPPAPGTDKSTPLTNKQPPGVDSLSATEALGPLSNAMVLPQAPMPRK 938
Db 834 TSGTTPSEM--YLPVPFSSESTRVRRG--AKSPED-----GPSAR-----QPLPRR 876
QY 939 SOATKLKPKRVKALYNCVADNPDELTFSEGDIIVDGEDQEWIIGHIDGDPGRKAPPV 998
Db 877 NVP-----VITEGDSGRTGSLPA 895
QY 999 SFVHFIAID 1006
Db 896 SSVQLQD 903

```

```

RESULT 11
Q7QY3 PRELIMINARY; PRT; 1015 AA.
AC Q7QY3
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE AGCP3317.
GN Name=agCG57483; ORFNames=ENSGANGG00000018249;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

```

```

CC Preliminary data.
DR EMBL; AA0801008807; EAA04701.1; -.
DR HSSP; Q60631; IGBQ.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3.1; 1.
DR PRINTS; PR00405; REVINTRACTNG.
DR PRODOM; PD000066; SH3; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 1015 AA; 111957 MW; DE8972F0E6CC898C CRC64;

Query Match 28.4%; Score 1499.5; DB 2; Length 1015;
Best Local Similarity 32.5%; Pred. No. 2.5e-55;
Matches 387; Conservative 150; Mismatches 294; Indels 359; Gaps 25;

QY 1 MPDOISVSERFAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLKMKSKVA 60
Db 1 MPELIGVEFEETRETDYNSFTTFTVSRMAQCRQTATLEETLDFDREGLTKLKAKVA 60
QY 61 INSSGLAHVNEEQYQTALEKFGGNCVCRDDPDLSGSAFKFSVFTKELTALFNLIQNMN 120
Db 61 IHNSGNTHVDNEMCLVRALERLGSVALSKEEPIGAAFLKFSVVTKELLSALMKTLMQIN 120
QY 121 NIISFPLDLSLKGDLGVKGLKPPFKAMKDYETKITIEKKEKHAHLKGMITEISG 180
Db 121 NIIVFPVDSLKSLRGKMGEMKPPDKAAKDYDKLMIKIEKKALAKEVGMRTVTP 180
QY 181 AEIAEEMEKRRFQOLQMCVLLKNEIKIKGVDDLQNLIKYFHAQCNFPDQGLKAVES 240
Db 181 AEIAEIEKERRVQLOMCCEYLIFKNEIKTKKGLLQHLVEYHQAQNNYFQGLUKTAH 240
QY 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLILKASAL---OVEQKEDSIQIRSTAYSLHQ 297
Db 241 FGTYIELSIKLQIRHKQDEERRKLELRLTLLSTPDPDRVENVPSCDKSCAIGYSLHQ 300
QY 298 PQGNKHGTERNGSLYKKSQDGIKRWQKRKCSVKNQGLTISHGTANRPAPKLNLLTCQVK 357
Db 301 LQGDKNHD----- 308
QY 358 TNPEKKCFDLISHDRTYTHFOADEQEQIWMVSLQNSKEEALNNAFKGDNDTGENNIVQ 417
Db 309 -----NRSYHFOADEADQKAWMSVLINCKEKALAKAFQAHANPQMSPLI- 353
QY 418 ELTKETIISVQRTMGNDVCCDGA-PDPTWLNGLTICIECSGIRHRELGVHSPMOSL 476
Db 354 ELQTVIKHQLNLPNDQCCDGRNDVTWLSLNFGLVLCIQCSGVHRLDGVHHSRIQSL 413
QY 477 TLDVLGTSELLAKNIGNAGNEIMECCLPADSCKPNPMDGMARKDYITAKYIERYA 536
Db 414 TLDNLTTAQLLVARAMGNALNEVEATLA---QSKLTPESTMEERYDFIRAKYVAKYV 470
QY 537 RKXHADNAKLSHCEAVKTRDIFGLLOAYAGVDLTETKIPLANGHEPDETALHLAVRSV 596
Db 471 MRTCADDRDLRNDLEQAVINADLQQLQVWABGADLTCLVLP----- 511
QY 597 DRTSLHIVDFLVQNSGNDKOTGKSTALHYCCLTIDNAECLKLLLRGKASIEANESGET 656
Db 512 -----SSIEHAKMKREKS 523
QY 657 PLDIAXRLKHECHCELTQALSGRPNSHVYEWRLHEDLDESDDMDKELQ 711
Db 524 APD-----HINTOWN-IHDDGSTDFSDDDTALMDER--- 553
QY 712 SENRDRPISF-----YOLCSNQLQNSAVSLARDAANL-----AKE 748
Db 554 ---KSRGRPPFVGSDSPVALRSRSTCDISQSGSSPSSSCNPNIARDQRIPIPSGT 610

```

```
QY 749 KORAFMPSILQNETYCALLS-----GSPPPAQPAASSTTSAPPLPRNVCKVQTA 798
Db 611 SPKQNTSTVFSHVASGTSSVQNSIGMIGVSAFTSGSSPIGNSSSSVQGVSTSQUT 670
QY 799 SSA-----NTLWKTNSVSDGSRQSSQSDPPAV----- 827
Db 671 SSSVATFNTSTNKKFSSVNV-GSLKRTAPAPPPTTYGTLPHAPRHSQNLDELFGVSSQ 729
QY 828 -----HP-----PLPPLRVTSNPLTPPTPPPVAKTPSVMEALSQPSKPAP----- 868
Db 730 HVLHGLHHHPDMYSTLPHLRGSDPISNNSASSVPAALLSSEMRGGSTKSMQAQLLFDNK 789
QY 869 -----PGISQIRPP--PLPQPSPRLPQKPGARTDKSTPLTN--KGQ 907
Db 790 IPDREYPTSSGAASVAGAKLVLPAGEIIPOLKPVSNRPKQIPIPSQIAVTOISNTEKSF 849
QY 908 PRGPVDLSATBALGPLSNAMVLQPPAPMPKRSQA-----TKLKP----- 946
Db 850 ANG-----QSNESLTSVDDALIMRRKIKGPASSAGVDSNLDGDTKETSYSNYDTSGIRNA 905
QY 947 -----KRVKALYNVCVAD 958
Db 906 LDNSVASSGGHMSLNSYSRNDNSGDASCVDLDSFNLSSRSFGGAMRCRALYDCNAD 965
QY 959 NPDELTTFSEGDVLIIDGGE--EDQEWIIGHIDGDPGRKGAFPVSVFVHTIAD 1006
Db 966 NDELEFKEGEILIVINERTDDENWMEQIEGDSMRGMPFPVSVFQMMDD 1015

RESULT 12
Q6AWJ6 PRELIMINARY; PRT; 935 AA.
AC Q6AWJ6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE LP17217p (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; BT015252; AAT94481.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IBA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3 2.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTING.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
```

```
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain.
FT NON_TER 1
SQ SEQUENCE 935 AA; 102696 MW; 45BF94ED72BF0FC3 CRC64;

Query Match 26.2%; Score 1380; DB 2; Length 935;
Best Local Similarity 37.6%; Pred. No. 2.5e-50;
Matches 362; Conservative 127; Mismatches 259; Indels 174; Gaps 31;

QY 192 RFFOLQCEYLLKYNEIKKIGVLLQNLQIKYFHAQCQNFQDGLKAVESLXPSIETLSTD 251
Db 1 RLYQLQCEYLLKYKDKITKTGIELLQHLIEYHALSNYFKQGLQTIHFQTYICDLSK 60
QY 252 LHTIQQAQDEERRQLIQLRDILKSALQVEQKED---SQIROSTA-YSLHQPGQKNEHETE 307
Db 61 LHEIKQKQDEDRRSLLDLRTVLRSTPDPERVDNVPSSESRSRGAGYSLHQDQKHGVT 120
QY 308 RNSGLYKSDG-IRKVNQKRCVS-KNGFLTIISHTANRPPAKNLNLTTCVKTNPKEKK 365
Db 121 RQGHLLKKSEKGRVNVQKRCRVTSDGFLDIFHADESKPPTRVNLLTCQIKPVPDDKRG 180
QY 366 FDLISHDRTYHFOAEDEQCOIWMVLSQNSKEEALNNAFKGDNDTGENNIVQELTKEIS 425
Db 181 FDLISYRPHYHFOAEDEGDQKAWAVLVNCKEALTKAFQHANPQMSPSLV-ELQKTIVR 239
QY 426 EVQRTGNDVCCDCA-PDPTWLSNLGILFCIFCSGIHRELGVHYHSPMSLTLDLVQTS 484
Db 240 YVQLLPGNDRCCDCGRNDVTWISLNFILVICQSGVHRDLGVHHSRIQSLTLDLNTTA 299
QY 485 ELLAKNIGNAGFNEIMECCLPAEDSVKPNPGSDMNARKOYITAKYIERRYARKKHADNA 544
Db 300 NLLTARANGNSTLNDIMEAKL---GRGKLOHESMEERYDFIRAKYAKRVYMRCTSDDN 356
QY 545 AKLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLANGHEPDEPTALHLAVRSVDRTSLHV 604
Db 357 DLRCLEQAVNVADMSQLQVWAGADLTCLCP---SSDAGETALHLAVLREMSTLHV 413
QY 605 DFLVQN-----SGNLDKQTKGSGTALHYCCLTDNAECLKLLLRKASIEIANE 652
Db 414 DFLIQNPPKGLNKATNPAGLLD-VTGGK-NTALHLCALHRRRCWKLLLRSGADYELKNS 471
QY 653 SGETPLDIKELKHEKCEELTQALSGRFNSHVHVEYEWRLLED--LDESDDD--WDEK 708
Db 472 QNKTALDIAKEMGHNSCRELIECAIKREKSAFDHINTDNLNPNSEDGTFDSDETVIDER 531
QY 709 LQPS-----ENRRDRP-----ISFYQ-LGS-----N 729
Db 532 SSSSPIANCPSRQFTLPSGLPSYTHSAGTSPKHISVGQYLGSAITNVGNGPENGSSPS 591
QY 730 QLQSNVSLARDAANLAKEKORAFMPSILQNETYCALLS-----PPPAQ 775
Db 592 SASQSQVRAARNSLN-----MQSDLGHVTVGARKSTSTANWNSLKRTAPAPPPGT 642
QY 776 PAAPSTTS---APPLPRNVCKVQVTASANTLMKTNVSU-----DGG 815
Db 643 LGSASSSSFYGTLPHPPRHSQNFDAIDIRAHNNKQSLDVAYGTLPHLRSVSSPRGGGG 702
QY 816 SRQRSSSDP---PAVHPPLPLRVTSNPLTPPTPP-----PPVAKTPSV 856
Db 703 YGYGVSDPQGGSGNGSNNSLNPAMTTFCHKKSPSGESLNRNIHLAGAKLVLPPTGELPTL 762
QY 857 ----MEALSQPSKAPPAGISQIRPPPL-----PPQPSRLPQKKPA 893
Db 763 KHVDSKALTREKIP-PPG-----PPSREISNGQSNESSISSMDEGVPVAPRKLNVQS-A 814
QY 894 PGTDKSTPLTNKGQPRGPVDLSATEALGPLSNAMVLQPPAPMPKRSQATKLKPK----- 947
Db 815 NFPDYESWHTDMDSGGGLDHSASNSVSSSDNRNLSNP-DNPSKTKGAGLGKGFHYNGQ 873
QY 948 -RVKALYNVCVADNPDELTFSEGDVLIIDGGE--EDQEWIIGHIDGDPGRKGAFPVSVFVHI 1004
```

```
Db 874 RRCALYDCVADNDELEKFEKEGELIVLNERTDDENWMEGIEGQPTKRGKFPVSVFVHML 933
Qy 1005 AD 1006
Db 934 PD 935

RESULT 13
Q8N282
ID Q8N282 PRELIMINARY; PRT; 208 AA.
AC Q8N282;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ33802.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tongue;
RX Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Teraashima Y., Suzuki O.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yanazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hibiji T., Kobatake N., Inagaki H., Ikeda Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AK091121; BAC03588.1; --
DR HSSP; Q9N2M3; IUFF.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3.1; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 208 AA; 21869 MW; 8ABA541403C99B6F CRC64;
```

Query Match 20.4%; Score 1075.5; DB 2; Length 208;  
Best Local Similarity 81.4%; Pred. No. 2.8e-38;  
Matches 206; Conservative 1; Mismatches 1; Indels 45; Gaps 1;

```
Qy 754 MPSILQNETYCALLSGSPPPAQAAPSTTSAPPLPPRNVGKVQTASSANTLWKTNSVSD 813
Db 1 MPSILQNETYCALLSGSPPPAQAAPSTTSAPPLPPRNVGK----- 41

814 GGSQRSSDDPAVHPPLPLRVSTNPLTPTPPPPVAKTFSVMEALSQSPKPPPGISQ 873
Db 42 -----DPLTPTPPPPVAKTFSVMEALSQSPKPPPGISQ 75

874 IRPPPLPPQPSRLPQKKPAPGTDKSTPLTNKGQPRGPDLSATEALGPLSNAMVLP 933
Db 76 IRPPPLPPQPSRLPQKKPAPGADKSTPLTNKGQPRGPDLSATEALGPLSNAMVLP 135

934 PMPRKSQATKLKPRVRYALNCVADNPDELTFSGDVIIVDGEDEQEWIHDGDPGRK 993
Db 136 PMPRKSQATKLKPRVRYALNCVADNPDELTFSGDVIIVDGEDEQEWIHDGDPGRK 195

994 GAPPVSVFHFAD 1006
Db 196 GAPPVSVFHFAD 208

RESULT 14
Q6DEH1
ID Q6DEH1 PRELIMINARY; PRT; 307 AA.
AC Q6DEH1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein (Fragment)
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins P.S., Wagner L., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buotow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki A.I., Skalska U., Smaluk J., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC077145; AAH77145.1; -.
KW Hypothetical protein.
FT NON TER 307 307
SQ SEQUENCE 307 AA; 35463 MW; 5BE7CD31B38F5FDC CRC64;
```

Query Match 16.7%; Score 878; DB 2; Length 307;

Best Local Similarity 54.8%; Pred. No. 9.6e-30;  
Matches 167; Conservative 70; Mismatches 64; Indels 4; Gaps 2;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMLVKMKSVKA 60

```
Db 1 MPEKISIDFVSLAKEDLSSFGSGFQSKMDCRSTVAALBESLEQDQMSLQKMKKIYKA 60
Qy 61 INSSGLAHVNEEQYTOALEFGGNCVCRDPDGLSAPLKFVSVTTELTAFLKLIQNMN 120
Db 61 IHNSGLAHVNEKEQYTEVLNGLSHLSQDNNEVSTGFELNLAFTREVTALFKNLVQNLN 120
Qy 121 NIISFPDLSLLKGLDKGVKGLKKPDKAWDKYTKIKIEKKEHAHKLHGMTRTEISG 180
Db 121 NIISFPDLSLLKGLDKGVKGLKKPDKAWDKYTKIKIEKKEHAHKLHGMTRTEISG 179
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVKNEIKKGVLDLLQNLIKYHPAQCNNFPQDGLKAVES 240
Db 180 -DTAEDMERERNFQLOMCEYLLKVKNEIKKGVLDLLQNLIKYHPAQCNNFPQDGLKAAEN 238
Qy 241 LKPIETLTDLHIKQADERRLOLRILKLSALQVEQKDSQIRQST--AYSUHQHOP 298
Db 239 LGPTFVEKLAASVHTVRHEQDEVKLSQLRSLRQLQVDQKEDFLNRKNSGHGYSIHOP 298
Qy 299 QGNKE 303
Db 299 QKXKK 303

RESULT 15
Q9V367 PRELIMINARY; PRT; 762 AA.
AC Q9V367;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE CG30372-PA.
GN ORNames=CG30372;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.B.,
RA Abiril J.F., Aghayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan J.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
```

```
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminiker J.S., Bergman C.M., Krommiller B., Carlson J., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5];
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6];
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AE003838; AAF59133.3; -.
DR HSSP; Q60631; LGBQ
DR FlyBase; FBgn0050372; CG30372.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF04412; ArfGAP; 1.
DR Pfam; PF00109; SH3_1; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK REPEAT; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50115; ArfGAP; 1.
DR PROSITE; PS50002; SH3; 1.
DR ANK repeat; SH3 domain.
KW ANK repeat; SH3 domain.
SQ SEQUENCE 762 AA; 82178 MW; 9F0236BA1DA9B7FB CRC64;

Query Match 16.0%; Score 844; DB 2; Length 762;
```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:28:06 ; Search time 10576 Seconds

(without alignments)  
4609.110 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAPVSVFVFIAD 1006

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool\_h/US09914042/runat\_04082005\_090632\_2111/app.query.fasta\_1.1159  
-DB=genEmbl -QWRT=fastcap -SUFFIX=rgc -MINMATCH=0.1 -LOOPFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042 @CGN 1 1 6279 @runat\_04082005\_090632\_2111 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.hcg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sw.\*

13: gb.un.\*

14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5253	99.6	5711	9	AB007860 Homo sapi
2	4968.5	94.2	5544	9	BC063308 Homo sapi
3	4727	89.6	4713	10	AK172944 Mus muscu
4	4215.5	79.9	3989	5	BC070750 Xenopus l

5	3877.5	73.5	5072	10	BC080847
6	3874	73.5	2949	6	AR243296
7	3874	73.5	5954	6	AR243295
8	3691	70.0	4486	6	I74317
9	3417	64.8	4328	6	I74314
10	3155	59.8	2712	6	AR243298
11	3155	59.8	4595	6	AR243297
12	3018	57.2	4434	10	BC048818
13	3013.5	57.1	5330	4	AF112886
14	3013.5	57.1	5330	4	AF112886
15	3011	57.1	4150	10	AF075462
16	3000.5	56.9	4370	10	AF075461
17	2954.5	56.0	3456	6	AR243294
18	2954.5	56.0	4382	6	AR243293
19	2888.5	54.8	2109	6	CQ720712
20	2859.5	54.2	4148	10	AK122477
21	2465.5	46.8	4870	9	HSM805854
22	2410	45.7	5475	9	AB033075
23	2224	42.2	2583	6	CQ716704
24	2089.5	39.6	3812	6	AX747456
25	2089.5	39.6	3812	9	AK092291
26	2089.5	39.6	4050	6	AX741097
27	2089.5	39.6	4050	9	AB056749
28	2085.5	39.6	4094	9	BC060786
29	1922	36.4	1721	5	CR387820
30	1803.5	34.2	3830	6	CQ575946
31	1583	30.0	6688	6	CQ575945
32	1583	30.0	167107	3	AC093104
33	1583	30.0	226833	3	AE003838
34	1555	29.5	157559	2	AC020344
35	1433	27.2	132171	3	AC008370
36	1380	26.2	3808	3	BT015252
37	1307.5	24.8	3123	6	AX746789
38	1307.5	24.8	3123	9	AK091121
39	1301.5	24.7	858	5	CR385625
40	1139	21.6	1777	9	AK000253
41	1029	19.5	884	5	CR523567
42	878	16.7	1185	5	BC077145
43	731	13.9	2400	10	MMU92478
44	716	13.6	706	5	BX950804
45	695	13.2	563	6	AX261503

ALIGNMENTS

RESULT 1	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	linear	PRI 10-JAN-2004
LOCUS	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
DEFINITION	AB007860				
ACCESSION	AB007860				
VERSION	AB007860.1	GI:2662080			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE					
AUTHORS					
TITLE					

1 Ishikawa, K., Nagase, T., Nakajima, D., Seki, N., Ohira, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. Prediction of the coding sequences of unidentified human genes. VIII. 78 new cDNA clones from brain which code for large proteins in vitro  
DNA Res. 4 (5), 307-313 (1997)  
9811655  
9455477  
2 (bases 1 to 5711)  
Ohara, O.  
Direct Submission  
Submitted (06-OCT-1997) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kiearazu, Chiba 292-0812, Japan (E-mail:cdna@info.kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)  
Location/Qualifiers

## source

```

1. 5711
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG1091"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. 5711
/gene="KIAA0400"
<293...3361
/gene="KIAA0400"
/note="Start codon is not identified"
/codon_start=1
/protein_id="BAA23696.2"
/db_xref="GI:40788243"

```

## gene

## CDS

```

/translation="GGGAPAAVQRPRAEAMPDQISVSEFVAETHEDYKAPTASSFTT
RTACRNTVAIAEALDVRMVLVYMKKSVKAINSSGLAHVENESEQYQALKEKFGNC
VCRDPPDLGSAFLKPSVFTKELTALFKNLIQNMNIIISPLDLSLLKGLKGVKGLKK
PFDKAWDYETKTKIEKEKEHAKLHGMITEISGAIEAEMEKERRFPQLOMCEVL
LKNEIKIKGVLDLQNLKYPHQAQCNFFQDGLKAVESLKPSIETLSLDLHIIKQAD
BERQLQRLDILKALQVEQEDSIQKSTAYSLHQPQGNKEHGTENRSGFLYKKSOG
IRKWKQKKSQVKGFLTISHGTANRPAPKLLLTQVKTNPBKCKCFDLISHDRTH
PQADEQPCQIWMQSVLQNSKEBALNNAFKHGDNTGNNI VOELTKEIISVQRMGTND
VCCDGPDPWTLSNIGLITCECSGIHRELGVHYSRMQSLTLDVLGTSLLAKNI
CAAFNETMECLPAEDSVKPNQSDMNARKDYITAKYIERVARKHADNAKLHSL
QVGNLDRIFGLQAYADGVDLTEKIPLANGHEDEDTALHLAVRSVDRKTSUHVDFLV
QNSKNDRTQKGSSTALHYCCLDINAECILKLLRLKASIEIANESGETPLDIAKLKH
EHCSELTLQALSNSRSHVHYEWRLLEDDESDDDMEKQLQPSNNREDRPISEY
QLGNSQLQNSAVSLARDANLAKKQRAFMPDSILQNETYGALLSGSPPLPAQPAVSTT
SAPPLPPNNVKQVOTASSANTLWKTNSVSDGSGRORSSSDPPAVHPPLPLPLVSTN
PLTPTPPNVAKTSPVMEALSQPSKAPPGI SQIRPPPLPQPPSRLLPKQKPPAGADK
STPLTNKQQRGPVDLSATEALGPLSNAMVLQPPAPMPKSKQATKPKRVKALYNCV
ADNPDELTFSGSDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVPSFVHFITAD"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 3e-187 Length: 5711
Score: 2523.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 9 Gaps: 0

```

US-09-914-042-1 (1-1006) x AB007860 (1-5711)

```

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAla 20
Db 341 ATCCGGACACAGATCTCCGTGTCGGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG 400
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGCTCCAGCTTACCACCCGCGACGGCGAGTCCGGAACACTGTGCGCGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60
Db 461 GAGGAGGCTTTGGACGTGGACCGGATGGTTCTTTTCAAAATGAAGAAATTCGTGGAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGACGAGTACACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGTCAGAGATGACCCAGATTTTAGGAAGTGCCTTCCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTCACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAGATATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
Db 701 AACATAATCTCCTTCCCTTTGACACAGTTTGTCTGAAGGGGGACCTGAAAGGAGTGAAGGG 760

```

```

141 AspLeuLysLysProPheAspLysAlaTrpIysEspTyrGluThrLysIleThrIysIle 160
761 GATCTGAAAAGCCCTTTTGATAAAAGCTTGGAAAGACTATGAAACAAAATAAACCAAGATA 820
161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
821 GAAAAGGAGAAAAGGAAACACGCCAAGCTCCATCGGATGATTCGGACTGAATAAAGCGGA 880
181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
881 CGCGAAATTCGGAAGAGATGGAAGAGAGAGAGCGCTTCTTCCAGCTACAGATGTGCGAG 940
201 TyrLeuLeuLysValAsnGluIleLysIleLysGlyValAspLeuLeuGlnAsnLeu 220
941 TATCTGCTGAAGGTCAACAGAAATCAAGATTAAGAGGAGTAGATTTTACTTCAGAACTG 1000
221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
1001 ATCAATACITTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAAGCCGTGGAAGC 1060
241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
1061 CTCAACCTTCCATTTGAAACGCTGCTACGGATCTTACACGATCTTCAACAGCAAGCCAGGAT 1120
261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
1121 GAAGAAAGAGGCGAGTGTATACAGCTTCGAGATATTTTGAATCCGATTCGAGGTGAA 1180
281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
1181 CAGAAAGGAGACTCCCAAAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGA 1240
301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
1241 AACAAAGAACATCGGACCGGACCGGAGCGCTCTACAAAGAGAGTGAAGGATCCGA 1300
321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
1301 AAGTGTGGCAGAAAAGAAATGTTTCAAGTTAAATGGTTTTCTGACCATATCCATGGT 1360
341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCT 1420
361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
1421 GAGGAGAAGAGAGTGTCTTACCTTATTTTACATGACAGAACTTACCACCTTTTCAAGCTGAA 1480
381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu 400
1481 GATGAACAGGAATGTCAAAATATGGATGTCGTGCTGCANAAATAGCAAGAAGAGCTTTA 1540
401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
1541 AACATGCAITTAAGGGGAGTACAAATCTGGAGAAAATAACATCGTCCCAAGAACTGACA 1600
421 LysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
1601 AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGCTGTGCTGACTGTGG 1660
441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysValIleGluCysSer 460
1661 CGCCAGATCTCATGCTTTCACCAACCTGGGCATCTTCAGCTGCATGCATGCAGTGTTC 1720
461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
1721 GGAATCCACCGAGAGCTGGGGTTCATTTATTCAGGATGACAGTCCCTGACCTTAGATGTA 1780
481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
1781 CTGGGAACATCTGAGCTGCTGCTGCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 1840

```

QY 501 MetGluCysCysLeuProAlaGluSpSerValLysProAsnProGlySerAspMetAsn 520  
 Db 1841 ATGGAATGTGCTACCGAGCTGAGACTCAGTCAAAACCCCAAGCCAGGAGGACATGAAT 1900  
 QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
 Db 1901 GCAAGAGGAGTACATCACAGCCAAAGTACATCGAGAGAGATACCGAAGAGAGGAC 1960  
 QY 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
 Db 1961 GCGGATAACGGCGGAGCTTACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTT 2020  
 QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
 Db 2021 GGATTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAAAATCCCACTGGCCAA 2080  
 QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
 Db 2081 GGACATGAGCCGGATGAACCGCCCTCCACTTTCAGATTCAGATCCGTGGATCGAACCTCT 2140  
 QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
 Db 2141 CTTCACATGTAGACTTTTATGTTAGTTCAGAAACAGTGGGAACCTGGATAAACACAGAGGAAA 2200  
 QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
 Db 2201 GGCAGCAGAGCCCTGCACCTACTGCTGCTGCCCTCCAGCAATGCGGATGCTCAAGTTGCTC 2260  
 QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
 Db 2261 CTGGGGGAAGGCTTCATCGATAGCAACAGTCCAGGAGACTCCCGCTGGACAT 2320  
 QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
 Db 2321 GCCAAGCGCTCAGCAGCAGCAGCTGTGAGAGCTGTGACCAAGCTTATCTGGAGA 2380  
 QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
 Db 2381 TTTAATTTCTACGTTTACGTTTGAATATGATGGCGACTACTCCACGAAGACTGGATGAA 2440  
 QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
 Db 2441 AGTGATGACGATGAGTGAAGATTTGAGGCGCCAGTCCCAACCGCGGGAAGACCGGCC 2500  
 QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
 Db 2501 ATCAGCTTCTACAGCTGGGTCCAAACAGCTTACGTACAGCTGTATCTTTGGCCAGA 2560  
 QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
 Db 2561 GATGCTGCAACCTTGCCAGGAGAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAT 2620  
 QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
 Db 2621 GAGACTTACGAGGCGCTCTCAGTGGCAGCCACCTCCCGCCAGCTTGCAGCGCCCGC 2680  
 QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
 Db 2681 ACCACGAGCGCCCGCCCTTCCACGGAATGTGGCAAGTTCAGACAGCTTCCTCT 2740  
 QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
 Db 2741 GCTAACCCCTGTGGAAGACAACTCTGTAAGTGTGACCGTGGAGCGGCGGAGCTCT 2800  
 QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840  
 Db 2801 TCGTCAGATCCGAGCTGTCCATCCAGCTGCCCTCTTTCGCGTGACATCTACCAAT 2860  
 QY 841 ProLeuThrProThrProProValAlaLysThrProSerValMetGluAlaLeu 860  
 Db 2861 CCCCTGACCCCGAGCGCCCGCCCTGTCAGAGAGCGCCAGCGTAATGGAAGCTTGT 2920  
 QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880

Db 2921 AGCCAGCGGAGCAAGCCTGCCCGCCTGGGATCTCACAGATCAGCGCCCGCCCTCTGCC 2980  
 QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900  
 Db 2981 CCACAGCGCCCGAGCGCTCCCGCAGAGAGAGCTGCGCCGGGGCTGACAGATGCCACC 3040  
 QY 901 ProLeuThrAsnLysGlyGlnProProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920  
 Db 3041 CCACTGACCAACAAAGGCCAACCGAGAGAGCTGTGGATCTCTCTGCAACGGAAGCTCTG 3100  
 QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940  
 Db 3101 GGTCTCTGTCCAATGTATGTTCTGCAGCCCTGCACCATGCTTAGAAGTCGAG 3160  
 QY 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960  
 Db 3161 GCAACCAAGTTGAAGCTTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCC 3220  
 QY 961 AspGluLeuThrPheSerGluGlyAspValIleLeuValAspGlyGluGluAspGlnGlu 980  
 Db 3221 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCGTGGACGGGAGGAGGAGGAGGAG 3280  
 QY 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000  
 Db 3281 TGGTGGATGGCCACATTTGATGGAGATCTGTGTCGAAAGCGCATTCCTCCGCTGTCA 3340  
 QY 1001 ValHisPheIleAlaAsp 1006  
 Db 3341 GTGCACATTTATCGCTGAC 3358

RESULT 2  
 BC063308  
 LOCUS  
 DEFINITION Homo sapiens development and differentiation enhancing factor 2,  
 mRNA (cDNA clone MGC:71731 IMAGE:30345039), complete cds.  
 ACCESSION BC063308  
 VERSION BC063308.1 GI:38648786  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 5544)  
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McSwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywicki,M.I., Skalek,U., Smaluk,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 REFERENCE 2 (bases 1 to 5544)  
 AUTHORS Strausberg,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Stefan Hansson  
 cDNA Library Preparation: Michael Brownstein / Ted Usdin  
 Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 133 Row: h Column: 11  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4502248.

## FEATURES

## source

1..5544  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:71731 IMAGE:30345039"  
 /tissue\_type="Placenta, normal"  
 /clone\_lib="NIH MGC\_147"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

## gene

1..5544  
 /gene="DDEF2"  
 /note="synonym: KIAA0400, PAP, SHAG1, PAG3, Pap-alpha"  
 /db\_xref="LocusID:8853"  
 /db\_xref="MIM:603617"  
 294..3179

## CDS

/gene="DDEF2"  
 /codon\_start=1  
 /product="DDEF2 protein"  
 /protein\_id="AAH63308.1"  
 /db\_xref="GI:38648787"  
 /db\_xref="LocusID:8853"  
 /db\_xref="MIM:603617"  
 1..5544  
 /translation="MPDOIISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAALIEEAL  
 DVDRMLVYKMKSVKAINSGLAHVENEEQYQALEKEGNGRVCDDPDLSGLPKES  
 VFTKELTALFNLIQNMNIISFPDLSLGLKGVKGDLLKPPDKAWKDYETKITKI  
 EKEKHEAKHGMIRTEISGAIEAEEMEKERRFFQLOMCEYLLKVNKEIKKKVDLLQ  
 NLIIFYHAQCNFQDGLKAVESLKFSEITLTDLTIKQAQDEERRQLIQRDLKSA  
 LQVEKDSQIRQSTAYSLHQPQGNKEHGTNRGSLYKKSQGIKVKWKRCVKNGF  
 LTI SHGTANRPAPKLNLTQCVKNPBEKKCFDLISHDRTYVFOAEDBOEQIWNVSL  
 QNSKEALNNAFKGDDNTGNNIVQELTKEIISVQVMTGNDVCDGAPDPTWLSN  
 LGILTCIECSGIRHGLVHYHSMQSLTDVLTGTSSELLAKNIGNAGFNEIMECCLP  
 DSVKPNPQSDNWRKDYITAKYIERRYARKHADNAKLSLCEAVKTRDIFGLLQAY  
 ADGVDLTEKIPLANGHEPDETHLAVRSVDRSLHIVDFLVQNSGNDKQTKGSTA  
 LHVCLTNAECLEKLLKGAIEIANESGETPLDIAKRKHCEHEELLTQALSGRFN  
 SHVHEVEMRLHLLHDESDDDMDKLOPSNNREDRPI SFVLGSLNQLQSNVSLAR  
 DAANLAKQKAFWPSIILQNETYCALLSGSPPPQAAPSTTSAPPLPRNVGKDLPT  
 PTPPPVAKTQFVPSLQSPKAPPGISQRPPLPPQPPSPRQPPKAPGADKSTP  
 LTNKQPGRGVLDLATEALGPLSNMVLQPPAPMPKQSOAKTLKPRKRYALNCVADN  
 PDELTFSRGDVIIVDGEDQEWIIGHIDGDFGRGAFFVSVFVFIAD"  
 1209..1490  
 /gene="DDEF2"  
 /note="PH; Region: Plectretin homology domain"  
 /db\_xref="CDD:smart00233"  
 1554..1898  
 /gene="DDEF2"  
 /note="ArfGAP; Region: Putative GTPase activating protein  
 for Arf. Putative zinc fingers with GTPase activating  
 proteins (GAPs) towards the small GTPase, Arf. The GAP of  
 ARD1 stimulates GTPase hydrolysis for ARD1 but not ARFs"  
 /db\_xref="CDD:pfam01412"  
 2052..2312  
 /gene="DDEF2"  
 /note="ANK; Region: ankyrin repeats"

## misc\_feature

/db\_xref="CDD:cd00204"  
 2991..3170  
 /gene="DDEF2"  
 /note="SH3; Region: Src homology 3 domains"  
 /db\_xref="CDD:smart00326"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,19e-176 Length: 5544  
 Score: 4968.50 Matches: 955  
 Percent Similarity: 95.13% Conservative: 2  
 Best Local Similarity: 94.93% Mismatches: 4  
 Query Match: 94.23% Indels: 45  
 Gaps: 1  
 US-09-914-042-1 (1-1006) x BC063308 (1-5544)  
 Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlyeAla 20  
 Db 294 ATGCCGACCAGATCTCCGTGTCGGAATTCGTGCCGAGACCCATGAGGACTACAAGCGC 353  
 Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
 Db 354 CCCACGGCCTCCAGCTTCACCACCCGACGCGCAGTCCCGAAACACCTGTGGCGGCCATC 413  
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIlyeAla 60  
 Db 414 GAGGAGGCTTTGGACGTGGACCGGATGGTCTTTTACAAATGAAGAAATCCGTGAAAGCA 473  
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluTyrThrGlnAlaLeuGlu 80  
 Db 474 ATCAACAGCTCTGGCTGGCTCACGTGGAATGAAGAGCAGTACACCCAGGCTCTGGAG 533  
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheIlye 100  
 Db 534 AAGTTTGGCGCAACCGGTGTATGCAGAGATGACCCAGATTTAGGAAGTGCCTTCTGAAG 593  
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 Db 594 TTCTCAGTGTTCAGAGGAGTTGACAGACATTTTCAAAAACCTTGATTCAGATATGAAAC 653  
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValIlyeGly 140  
 Db 654 AACATAATCTCTCTTCCCTTTTGGACAGTTTGTGAAGGGGGACCTGAAAGGAGTGAAGGG 713  
 Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrIlye 160  
 Db 714 GATCTGAAAAAGCCCTTTTGATAAAGCTTTGGAAGGACATATGAACAAAAATAACCAAGATA 773  
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleAtgThrGluIleSerGly 180  
 Db 774 GAAAAGGAGAAAAGAAACACCCAGCTCCATGGGATGATTCGGACTGAATAAAGCCGA 833  
 Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
 Db 834 GCGGAAATTCGGAAGAGATGGAAGAGAGAGCGCTTCTTCCAGCTACAGATGTGCGAG 893  
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIlyeLysGlyValAspLeuLeuGlnAsnLeu 220  
 Db 894 TATCTGCTGAAGGTCAACGAAATCAAGATTTAAAAGGGAGTAGATTTTACTTCAGAAATCTG 953  
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
 Db 954 ATCAAAATCTTTCATGCCCCAATGCAATTTTTTTCAGGATGACTCAAGCCGTGGAAGC 1013  
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 Db 1014 CTCAAACTTCCATTGAAACGCTGTCTACGGATCTTTCACAGTCAACACAGGCCCCAGGAT 1073  
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
 Db 1074 GAAGAAAGAGGCGAGTTTGATACAGCTTCAGATATTTTGAATAATCCGATTCAGGTTGAA 1133

## misc\_feature

QY	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300		2214	CTGGGGGGAAGGCGCTCCATCGAGATAGCAACGAGTCCAGGAGACATCCGCTGGACAT	2273	
Db	1134	CAGAAAGAGGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGGA	1193					
QY	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320		661	AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg	680	
Db	1194	AACAAGGAACATGGAGCCGAGCGGAACCGGACGCTCTACAGAGAGGTGACGGATCCGA	1253		2274	GCCAAAGCCCTCAGCACGACGACTGTGAGAGGTGCTGACCCAGCCTTATCTCGAAGA	2333	
QY	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340		681	PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu	700	
Db	1254	AAAGTGTGGCAGAAAAGGAATGTTTCAGTAAATAATGGTTTTCTGACCATATCCCATGGT	1313		2334	TTTAATCTCAGCTTACGTTGAATATGAATGGCGACTACTCCAGGAAGACCTGGATGAA	2393	
QY	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360		701	SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro	720	
Db	1314	ACCGCTAACCGGCGCTCTGCAAGCTCAACCTGTAACTGCTGCAAGTGAAGACCAACCT	1373		2394	AGTGATGACGATGATGAGAAATTCAGAGCCCACTCCCAACCCGCGGGAAGACCGGCC	2453	
QY	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380		721	IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg	740	
Db	1374	GAGGAAGAAGTGTCTTTGACCTTATTTACATGATGACAGAACTTACCACTTTCAGCTGAA	1433		2454	ATCAGCTTCTACCAAGCTGGGTCCACACGACTTCTAGTCTACGCTGTATCTTTGGCCAGA	2513	
QY	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400		741	AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn	760	
Db	1434	GATGAACAGGAATGTCAATAATGATGATGCTGTGCTGCAAAATAGCAAGAAGACTTTA	1493		2514	GATGCTGCNAACCTTGCCAGAGCAAGCAGAGGGGCTTTTCATGCCCAGCATCTTCGAGAAT	2573	
QY	401	AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420		761	GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaIleProSer	780	
Db	1494	AACATGCAATTTAAGGGGATGACAATATCTGGAGAAATAACATCGTCCAAGAACTGACA	1553		2574	GAGACTTACGAGAGCCCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGCAGCCCCCAGC	2633	
QY	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440		781	ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer	800	
Db	1554	AAGGAGATCATCTCAGAAGTGAGAGGATGACGGCAATGACGTCTGCTGTGACTGTGGG	1613		2634	ACCACGAGCCCCCGCTCTCTCCACGGAAATGTTGGCAAA-----	2675	
QY	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460		801	AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer	820	
Db	1614	GGCCAGATCCTACATGCTTTCCACCAACCTGGGCATCCTGACCTGCAGTGTTC	1673		2675	-----	2675	
QY	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480		821	SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn	840	
Db	1674	GGAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGGATGCAGTCCCTGACCTTAGATGTA	1733		2676	-----	2676	
QY	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500		841	ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu	860	
Db	1734	CTGGGAACATCTGAGCTGCTGCCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC	1793		2679	CCCTTGACCCCAAGCCGCCCCACCCCTTGCCAAAGAGCCCGAGCTTAATGGAAGCCTTG	2738	
QY	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520		861	SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro	880	
Db	1794	ATGGAACTGTGCTTACCACTGAGACTCAGTCAACCCACCCAGGAGCGACATGAAT	1853		2739	AGCCAGCGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCACCTCTGCCC	2798	
QY	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHis	540		881	ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr	900	
Db	1854	GCAAGAAAGGACTACATCACAGCCCAAGTACATCCGAGAGAGATACGCAAGGAAGCAC	1913		2799	CCACAGCCCGCCAGCGCTCCCGCAGAAGAAGCCTGCGCGCGGGGCTGACAAGTCCACC	2858	
QY	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560		901	ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu	920	
Db	1914	GCGGATAACCGGCGGAAGCTTCACAGTCTTTGCGAGGCGCGTCAAAACGAGAGATATTTT	1973		2859	CCACTGACCAACAAGGCCAACCGAGAGAGACTGTGGATCTCTGTGCAACGGAAGCTGTG	2918	
QY	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580		921	GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln	940	
Db	1974	GGATTGCTCCAAAGCTTATGCTGATGGTGTGGATCTTACGGAAAAAATCCCACTGGCCAAC	2033		2919	GGTCTCTGTCCAATGCTATGCTCTGCTGAGCCCTCCACCCATCCCTAGGAAGTCGAG	2978	
QY	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600		941	AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro	960	
Db	2034	GGACATGAGCGGATGAAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT	2093		2979	GCAACCAAGTTGAAGCCTTAAGCGGGTGAAGCGCTCTATAACTGTGTGTGGCTGACAACCC	3038	
QY	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620		961	AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu	980	
Db	2094	CTTCATTTGTAGACTTTTATGTTGAGAACACAGTGGGAACCTGGATTAACACAGACGGGAA	2153		3039	GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGGAGCGGGAGAGACCCAGAG	3098	
QY	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640		981	TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe	1000	
Db	2154	GGCAGCAGCCCTGCATCTACTGCTGCTGACCGCAATGCGAGTGCTCAAGTTGCTC	2213		3099	TGGTGGATTGGCCACATTTGATGGAGATCTCTGTGTCGCAAGGCGCATTCCTCGGTGTCTTT	3158	
QY	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660		1001	ValHisPheIleAlaAsp 1006		

Db 3159 GTGCACCTTATCGCTGAC 3176  
RESULT 3  
AKI172944  
LOCUS AKI172944 4713 bp mRNA linear ROD 28-JUL-2004  
DEFINITION Mus musculus mRNA for mKIAA0400 protein.  
ACCESSION AKI172944  
VERSION AKI172944.1 GI:50510472  
KEYWORDS FLI\_CDNA.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 Okazaki, N., Kikuno, R.F., Ohara, R., Inamoto, S., Koseki, H.,  
Hiraoka, S., Saga, Y., Seino, S., Nishimura, M., Kaisho, T., Hoshino, K.,  
Kitamura, H., Nagase, T., Ohara, O. and Koga, H.  
Prediction of the Coding Sequences of Mouse Homologues of KIAA  
Gene: IV. The Complete Nucleotide Sequences of 500 Mouse  
KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences  
of cDNA Clones Randomly Sampled from Size-Fractionated Libraries  
DNA Res. 11, 205-218 (2004)  
2 (bases 1 to 4713)  
REFERENCE  
1 Okazaki, N., Kikuno, R.F., Nagase, T., Ohara, O. and Koga, H.  
Direct Submission  
Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research  
Institute, Laboratory for Genome Informatics; 2-6-7  
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)  
The CREATE program supported by Japan science and technology  
corporation; cDNA full insert sequencing: Kazusa DNA Research  
Institute; cDNA library construction, clone selection and 5'- &  
3'-end one pass sequencing.  
LOCATION/Qualifiers  
1. .4713  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="mfj04247"  
/note="vector:modified pBC SK+"  
1. .4713  
/gene="mKIAA0400"  
<65. .2977  
/gene="mKIAA0400"  
/note="CDS is predicted by in silico analysis. Start codon  
is not identified."  
/evidence=not experimental  
/product="mKIAA0400 protein"  
/protein\_id="BAD3222.1"  
/db\_xref="GI:50510473"  
/translation="GGGSGPRGVAAMPDQISVEFVAETHEDYKAPTASSFTTRTAQ  
CRNTVAATEALDVRMVLVYKMKSVKAINISGLAHVENEREQYTOALEKFGCNCVCRD  
DPDLGSAFLKFSVFTKELTALFKLIQNMNIIIPPLDSLKGLKGVKGLKKPFDK  
AWKDYETIKIEKEKBEKHLGMIRTEISGAIEABEMEKERRFFOLQWCEYLLKVN  
EIKVYKQDILLQNLKYPAOCNPFQDGLKAVESIKPSITLSTDLTIKQAQDEER  
QLIQRDLKALQVEKESRDSQLRSTAYSLHQPQCNKEHGTENGNLYKSDGI  
RKWQKRCVSWNGSLTISHGTANRPAPKLLTCCQVNTPEPKKCFDLISHDRTYHF  
QAEDEQECQIWMVQLNSKEELANNAFKGDNDTGNNTIVQELTEIISSEVQRMGNDV  
CCDGAEPDTWNLNLGLTICEGSIHRELGVHYSRMQSITLDVLGTSLELLAKNLIG  
NAGNEIMECCLPSEDPKPNPDSMIARKDYITAKVMERYARKKHADTAALKHSLC  
EAVKTRDIFGLQLQAVADGVDLTKETPLANGHEPDDETALHLAVRSVDRPSTLHIVPLVQ  
NSGNDLKQGTQSTALHYCCILTDNAECLKLLRGSKASIEIANESGETFLDIKRLKHE  
HCELLDTQALSGRNFVHVHVEYERLHLLDDEDDDDVDEKLQPSPNRDRPVSFYQ  
LGSSQFQVNSQADTANLTKQKRGFGPSILQNETYGAILSGSPSSQIPSPSTTS  
APPLPRNVAGKDLTTTPPPVAKTSGTLEAMNQPSSKSPQTSQSKPPPIPPPPSR  
LPQKPKASGDKPTPLTNKGQPGPEASGPISNAMAQLPPAPMPKRSQATSKPKRVK  
ALYNCVADNPDELIFSEGVDIIVDGEEDQEWIGHIDGEPSPKGAFFVSVFHFIAID"

Percent Similarity: 91.77% Conservative: 18  
Best Local Similarity: 89.99% Mismatches: 29  
Query Match: 89.65% Indels: 54  
DB: 10 Gaps: 3  
US-09-914-042-1 (1-1006) x AKI172944 (1-4713)  
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
Db 101 ATGCCGGACAGATCTCCGTGTCGGAATTCGTGCCGAGACCCATGAGGACTACAGCGC 160  
QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40  
Db 161 CCCAGGCGCTTAGCTTCCACCACGCGCAGCGCCAGTGCAGAACACCGTGGCGGCATC 220  
QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60  
Db 221 GAGGAGGCTTTGGACGTGGACCGGATGGTTCTCTACAAAATGAAAGAAATCTGTGAAGCA 280  
QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80  
Db 281 ATCAATATCTCTGGGCTGGGCACACGTGGAGAACGAGGAACAGTACACCCAGGCTCTGGAG 340  
QY 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
Db 341 AAGTTCCGTGGCACTCTGTGTGCAGAGATGATCCAGACTTAGGAGCGGCTTCTGAAG 400  
QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLleGlnAsnMetAsn 120  
Db 401 TTCTCTGTGTTTCAACAAGGAGTTGACGGCATCTCTTCAAAAACCTGATTCAGAACATGAAC 460  
QY 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 461 AACATCATCTCTGTTCCCTTTGGACAGTTTGTCTGAAGAGATCTGAAAGAGTAAAGGG 520  
QY 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 521 GATCTGAAAAGGCCCTTTGATAAAGCTTGGAAAGGACTATGAAACAAAATAAACCAATA 580  
QY 161 GluLysGlnLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180  
Db 581 GAAAGGAGAGAAAAGGAGCATGCCAAGCTCCACGGGATGATTCGTACTGAAATTAAGCGG 640  
QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
Db 641 GCTGAGATTGAGAGGAGATGGAGAGAAAGACGGTTCTTCCAGCTGCAGATGTGTGAG 700  
QY 201 TyrLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 701 TACCTGCTGAAGGTCAATGAAATCAAGGTGAAAAGGAGTGGATTGTCTCAGAACTCG 760  
QY 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240  
Db 761 ATCAAGTACTTTCACGCCCATGCAATTTTTTCAGGATGGATTTGAAAGCAGTAGTAAAGC 820  
QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 821 CTCAAGCGCTTCCATCGAGACGCTCTCCAGGACCTCCACACCATCAACACAGCCCGGAT 880  
QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 881 GAGGAACGGCGACAGCTGATACAACTTCGAGATATTTTGAAGTCAGCATTCGAGGTGAA 940  
QY 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
Db 941 CAGAAAGAGCTTAGGAGAGACTCAAACTTCGCCCAAGTACAGGCTTACAGTTTACATCAA 1000  
QY 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
Db 1001 CCTCAGGAAACAAGGAGCATGGACAGAGAGAACCGGGAACCTCTACAGAGAGCGAT 1060  
QY 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337

Alignment Scores:  
Pred. No.: 1.03e-167 Length: 4713  
Score: 4727.00 Matches: 908

ORIGIN



Db 1061 GGGATCCGGAAGTGTGGCAGAGAGGAAGTGTTCGGTTAAAAATGGCTTCTCTACCATC 1120  
 Qy 338 SerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
 Db 1121 TCCACGGCAGCTGCCAAACGGGCACCTGCGAAGCTCAACCTGCTAACCTGCCAGGTGAAG 1180  
 Qy 358 ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377  
 Db 1181 ACCAATCCGAGGAGGAAGGTTCGACCTCATATCAGTACAGAGACGTACCACCTTC 1240  
 Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGlu 397  
 Db 1241 CAGCGGAGAGCAACAGAGATGTCAGATATGGATGCTGTACTGCAGAACAGCAAGGAA 1300  
 Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417  
 Db 1301 GAAGCTCTGAAACAACGCCCTTTAAGGGTGATGACACACCTGGAGAAAAATAACATGTCCAA 1360  
 Qy 418 GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
 Db 1361 GAGCTGACCAAGGAGATCATCTCGAGGGTGCAGAGGATGACGGCAATGACGTGTGCTGC 1420  
 Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGluIleLeuThrCysIle 457  
 Db 1421 GACTGTGGGACACAGATCCGACGTGGCTCTCTACCAACCTGGGCATCTGACTTGCATC 1480  
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
 Db 1481 GAGTGCTCTGGATCCACGGGAGCTGGGGTTCATTACTCCAGGATGCAGTCCCTGACA 1540  
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
 Db 1541 TTAGATGTATTGGGAACCTCTGAGCTTCTGCTTGCCAAAGATATCGGGAATGACGGCTTT 1600  
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
 Db 1601 AATGAATATTGGAGTGTTGCTACCGTCTGAGGACCCAGCTCAAGCCAAACCCAGGCAGT 1660  
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537  
 Db 1661 GACATGATTGCAAGGAGAGCTACATCACAGCAAGTACATGGAGAGAGATACGCACGG 1720  
 Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
 Db 1721 AAAAAGCATGCAGACACCGCGCGAAGCTCCACAGCCTTTCGAGCGCTCAAGACCAGA 1780  
 Qy 558 AspilePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
 Db 1781 GACATTTTGGGTACTCCAGCTTATGCTGATGTGTGGACCTGACAGAGAAATCCCA 1840  
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
 Db 1841 CTGGCCNATGGCATGAGCAGATGAGACAGCCCTCCATCTTGAGTCAGATCTGTGGAC 1900  
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
 Db 1901 CGGACTTCCCTTCATATTGAGACTTCTGCTGCTCCAGAACAGTGGGAACCTCGATAAACAG 1960  
 Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeu 637  
 Db 1961 ACAGCAAGGGCAGCAGACAGCCCTGCACCTACTGCTGCTGCAGACAGCAATGCTGAGTGCCTC 2020  
 Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657  
 Db 2021 AAGCTCTCTCGGGGGAAGAGCTTCCATTGATAGCCATGAGTCAGGAGAGACCCCG 2080  
 Qy 658 LeuAspileAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuLeuThrGlnAlaLeu 677  
 Db 2081 TTGGACATTTGCTAAGCGCTCAAGCACCAACATCTGTAGGAACCTCTCACTCAGGCCCTTG 2140  
 Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697  
 Db 2141 TCGGGGCGGTTTAACCTCCACGTTTCACGTCGAGTATGAATGGCGCTCTGTCATGAGGAC 2200

Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGlu 717  
 Db 2201 CTGGATGAGAGCGACGAGATGTGGAGAGCTTACAGCTAGTCCCAACCGCGGAG 2260  
 Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737  
 Db 2261 GACCGGCTGTAGCTTCTACAGCTGGGGTCCAGCCAGTTTCAGTCAATGCTGTGCT 2320  
 Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757  
 Db 2321 TTGGCCAGAGATCTGCAAACTCCCAAGGACAGCAGAGGGGCTTTGGGCCAGCATC 2380  
 Qy 758 LeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAla 777  
 Db 2381 TTGCAGATGAGACCTATGGAGCCATCTGAGTGGCAGCCGCCCTCTCTCCAGTCAATA 2440  
 Qy 778 AlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThr 797  
 Db 2441 CCCCCAGCACCCAGCGCACCCCGCTCCACCTCGGAATGTTGGCAAA----- 2491  
 Qy 798 AlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArg 817  
 Db 2491 ----- 2491  
 Qy 818 GlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArgValThr 837  
 Db 2491 ----- 2491  
 Qy 838 SerThrAsnProLeuThrProThrProProProValAlaLysThrProSerValMet 857  
 Db 2492 -----GATCCCTGACCCACACACACCCCTCTCTGTGGCTAAGACACTCAGAACTCTA 2545  
 Qy 858 GluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProPro 877  
 Db 2546 GAAGCATGAACAGCGAGCAGTCTCCCAACCTGGGACCTCACAGAGCAAGCCCCCG 2605  
 Qy 878 ProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAsp 897  
 Db 2606 CCCCTGCCACCTCAGCCACCCAGCGCTGCTCAGAAAGAAACACAGTTCGGGGAGTGCAC 2665  
 Qy 898 LysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThr 917  
 Db 2666 AAGCCAAACCCGCTGACCAACAAAGGACAGCGAGAGGACCT----- 2707  
 Qy 918 GluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArg 937  
 Db 2708 GAAGCTTCGGGTCTCTGTCCAACGCCATGGCCCTGCGAGCCCCCGACCCATGCTAGG 2767  
 Qy 938 LysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAla 957  
 Db 2768 AAGTCACAGGCGCAAGTCAAGCCCAAGCGAGTGAAGACACTCTACAACTGCGTGGCT 2827  
 Qy 958 AspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGlu 977  
 Db 2828 GACAACCCAGACGAGCTGACTTCTCTGAGGGCGACGTCATCATTTGTGTGGTGGGAAGAG 2887  
 Qy 978 AspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhePro 997  
 Db 2888 GATCAGGAGTGTGGATCGGCCCATCATGATGAGAGCCCGACCGCCAAAGGAGCGTTTCCC 2947  
 Qy 998 ValSerPheValHisPheIleAlaAsp 1006  
 Db 2948 GTATCGTTTGGCACTTTATTGCTGAC 2974

## RESULT 4

BC070750

LOCUS

DEFINITION

Xenopus laevis hypothetical protein MGC83760, mRNA

ACCESSION

BC070750

VERSION

BC070750.1

KEYWORDS

MGC.

3989 bp mRNA linear VRT 19-JUL-2004  
 MGC:83760 IMAGE:6640628), complete cds.

SOURCE	Xenopus laevis (African clawed frog)	
ORGANISM	Xenopus laevis	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.	
REFERENCE	1 (bases 1 to 3989)	
AUTHORS	Klein,S.L., Strausberg,R.L., Wagner,L., Pontius,J., Clifton,S.W. and Richardson,P.	gene
TITLE	Genetic and genomic tools for Xenopus research: The NIH Xenopus Initiative	
JOURNAL	Dev. Dyn. 225 (4), 384-391 (2002)	
PUBMED	12454917	CDS
REFERENCE	2 (bases 1 to 3989)	
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusica,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettner,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
PUBMED	12477932	
REFERENCE	3 (bases 1 to 3989)	
AUTHORS	Klein,S. and Strausberg,R.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-MAY-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA	
REMARK	NIH-MGC Project	
COMMENT	Contact: XGC help desk Email: cgapsb@mail.nih.gov Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu, Parvaneh Saedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smailus, Jeff Stott, Miranda Teali, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Place: 158 Row: f Column: 6 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..3989 /organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:8355" /clone="MGC:83760 IMAGE:6640628" /tissue_type="Oocytes"	
source		

	/clone_lib="NICHD_XGC_001"	
	/lab_host="DH10B"	
	/note="vector: pCMV-SPORT6.1"	
	1..3989	
	/gene="MGC83760"	
	/db_xref="LocusID:431962"	
	369..3458	
	/gene="MGC83760"	
	/codon_start=1	
	/product="MGC83760 protein"	
	/protein_id="AAH70750.1"	
	/db_xref="GI:47123082"	
	/db_xref="LocusID:431962"	
	/translation="MPDLISVSEFVCTNEDYKPSATSNFTNRLSQCNVTVAIDEAL EVDHTILNKKKSVKALHTGLTHVNEEQVTOALEKFGNSCVCRDDPDGLTAEFLKPS EVEHTILNKKKSVKALHTGLTHVNEEQVTOALEKFGNSCVCRDDPDGLTAEFLKPS EKEKHAHAKHGMIRITISGABIAEENKERRVFOQMCEYLLKNIKIKKGVLLDQ NLIKYFHAQCNFFODGLKAVETLAPSIKLSITDLHTIKQVDEERKQLSQRLDKAA LQVREDSDIROSQSTAYSLHPOQNGKHGTEDGVLKYSKDGIRKWKQKCKTVNGF LTIHQGTANIPPAKLLTQVKTNPBKKSFOLIHDRTVYFOAEDEAEQCIWMSVL QNSKEALNNAFKDHSAGENNIVQELTKELISIQMPGNDVCCDGGASPTWLSTN LGILITCECSGHHRELGHVYSRIOSLITLDVLGTSELKLNKNGNSVNEVNEACLPAD DSLKPTPCSDNNARKDYITTYIERKRYARKYHDNASKLHGLCDVAKSRDIPALVHLH AEGVDLTAEPPLANGHEQGETALHVAVSVDNLSLHIVDFLVQNSGLDKQTTKGSTA LHYCCLTDNAECLKLLRGKASTDIANDAGETPLDIAKRLKHAQCEDLLIOAMNGFN SHVHEVEMRLHVDLDESODDDVDEKMQOSPISREERPVSFYOLGINOGPSVACLA REASIAKDFRPNLPIQNETYGVNIGNSNFCNAGAPATTNAPLPRTNTIKVHS TIFGSSIHITPSNTPAWKASSISAESRQKSSSDPFISIQPVPLRVTSTATPPFP TKGPPALVTASLPAADPFSETTSQSGPKPSKLTSPQPPAPMRKSLQPKMKPKRV KAIYHCSADNPDELTFSEGEIIVVDGEDEKEMWIGHIDGEPNRRGAFVPSVHFPIAE"	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.04e-148	Length: 3989
Score:	4215.50	Matches: 806
Percent Similarity:	86.42%	Conservative: 85
Best Local Similarity:	79.18%	Mismatches: 113
Query Match:	79.95%	Indels: 27
DB:	5	Gaps: 7
US-09-914-042-1 (1-1006) x BC070750 (1-3989)		
Qy	1	MetProbapGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db	369	ATGCCGGAATCTGATCAGTGTCGGAATTTGTGTCGAGACCAATGAGGATTCAGATCC 428
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40
Db	429	CCCACTGCCTCCAATTTCAACCAACAGGCTGAGCCAGTGCAGGAAACACGGTCTGTGCTATA 488
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db	489	GATGAGGCTCTCGAGTTGATCACACGATCTCTGAAACAAAATGAAAAATCTGTGTAAGGCC 548
Qy	61	IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyrThrGlnAlaLeuGlu 80
Db	549	ATTACACATCTTGGTCTGACCCATGTGAAATGAGGAGCAGTACACTCAAGCACTGGAA 608
Qy	81	LysPheGlyGlyAenCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db	609	AAATTCGGAAGCAACTGCGTTTGTAGGACACCTGATTTGGGAACGGCTTCCTCAAA 668
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuLeuGlnAenMetAen 120
Db	669	TTCTCTGTCTTCCACAGGAGCTAACTGCACCTTTTCAAAAAACCTGGTGCAGAACATGAAT 728
Qy	121	AenIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db	729	AACATCATCATCTTCCACTTTCACAGCTTACTTGAAGGAGATCTCAAGGGGGTGAAGGG 788
Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160





Query Match:	73.53%	Indels:	197
DB:	10	Gaps:	3
US-09-914-042-1 (1-1006) x BC080847 (1-5072)			
QY	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla	20
DB	289	ATGCCGGACCAAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTTACAAGCGC	348
QY	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
DB	349	CCACGGCCTTAGCTTCACCAACGGCAGCCAGTCCGCGAACACCGTGGCGGCATC	408
QY	41	GluGluAlaLeuAspValAspArgMetValIleuTyrIysMetLysLysSerValIysAla	60
DB	409	GAGGAGGCTTGGACGTGGACCGGATGTTCTCTACAAAATGAAGAAATCTGTGAAGCA	468
QY	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu	80
DB	469	ATCAATATCTCTGGGCTGGCACACGTGGAGAACGAGAACAGTACACCCAGGCTCTGGAG	528
QY	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys	100
DB	529	AAGTTCGGTGGCAACTGTGTGCAGAGATGATCCAGACTTAGGAGCGCGTTTCTGAAG	588
QY	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
DB	589	TTCTCTGTGTTCAACAGGAGTTGACGCACCTTTCAAAACCTGATTCAAGAACATGAAC	648
QY	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly	140
DB	649	AACATCATCTCGTTCCTTTCGACAGTTTGTGAAGGAGATCTGAAAGGAGTAAAGGG	708
QY	141	AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrIysIle	160
DB	709	GATCTGAAAAGCCCTTTGATTAAGCTTGGAGGACTATGAACAAAATAATACCAAAATA	768
QY	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
DB	769	GAAGAAGGAGAAAAGGAGCATGCCAAGCTCCACGGGATGATTCTGTACTGAAATAAGCGG	828
QY	181	AlaGluIleAlaGluLeuMetGluLysGluArgPhePheGlnLeuGlnMetCysGlu	200
DB	829	GCTGAGATTGAGAGGAGATGGAGAGGAGAAAGACGGTTCTTCCAGCTGCAGATGTGTAG	888
QY	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu	220
DB	889	TACCTGCTGAAGTCAATGAATCAAGGTGAAGGAGGAGTGGATTGCTTCAGNAATCTG	948
QY	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
DB	949	ATCAAGTACTTTCACGCCCCAGTGCATTTTTCACAGGATGGATTGAAAGCAGTAGAAAGC	1008
QY	241	LeuLysProSerIleGluThrIleuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
DB	1009	CTCAAGCCTTCCATCGAGACGCTCTCCACGGACCTCCACACCATCAAAACAGGCCAGGAT	1068
QY	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
DB	1069	GAGGACCGGCACACTGATCAACTTCGAGATATTTTGAAGTCAGCATTTGAGGTGGAA	1128
QY	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
DB	1129	CAGAAAGAGGACTCACAACTTCGCCAAAGTAGCAGCTACAGCTTACATCAACCTCAGGGA	1188
QY	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
DB	1189	AACAAGGAGCATGGAACACAGAGGAAACCGGAACCTCTCAAGAACAGAGCATGCGATCCG	1248
QY	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
DB	1249	AAAGTGTGGCAGAGAGGAGTGTTCCTGTTAAATAATGGCTTCCTCACCATCTCCACCGC	1308

QY	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
DB	1309	ACT-----	1311
QY	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
DB	1311	-----	1311
QY	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu	400
DB	1311	-----	1311
QY	401	AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
DB	1311	-----	1311
QY	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
DB	1311	-----	1311
QY	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
DB	1311	-----	1311
QY	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
DB	1311	-----	1311
QY	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
DB	1312	-----CTTGCCCAAGAAATATCGGAAATGCGAGCTTTAATGAAATTT	1350
QY	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
DB	1351	ATGGAGTGTTCCTACCGTCTGAGGACCCAGCTCAAGCCCAACCCAGGAGTACATGATT	1410
QY	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis	540
DB	1411	GCAAGGAAGGACTACATCACAGCAAGTACATGGAGAGAGATATGCACGGAAGAAAGCAT	1470
QY	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560
DB	1471	GCCACACCGCGGGAAGCTCCACAGCCTTTGCGAGGCCGTCAGAGCCAGACATTTT	1530
QY	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
DB	1531	GGGTACTTCCAAGCTTATGCTGATGGTGGACCTGCACAGAGAAATCCCACTGGCCAAT	1590
QY	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
DB	1591	GGGCATGAGCCAGATGAGACAGCCCTCCATCTTCAGTCAGATCTGTGGACCGGACTTCC	1650
QY	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620
DB	1651	CTTCACATTGTAGACTTCTCGTCCAGAACAGTGGGAACCTGGATAAACACAGACGCAAG	1710
QY	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
DB	1711	GGCAGCACAGCCCTGCATCTGCTGCCTGACAGACAAATGCTGAGTGCCTCAAGCTCCTC	1770
QY	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
DB	1771	CTCGGGGAAAGGCTCCATTGAGATAGCAATAGTCCAGGAGACACCCGTTGGACATT	1830
QY	661	AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg	680
DB	1831	GCTAAGCGGCTCAAGCAGCAACACTGTGAGAACTGCTCACTCAGGCTTGTGGGGCGG	1890
QY	681	PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu	700
DB	1891	TTTAACTCCCATGTTACGTCAGTATGAATGGCGGCTGCTGCACAGGAGACTGGATGAG	1950
QY	701	SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro	720

```

Db 1951 AGCGACGACGATGTGGAGAGAGCTACAGCTAGTCCCAACCGACGCGGAGCGGCT 2010
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2011 GTTAGCTTCTACAGCTGGGTGCCAGCAGTTTTCAGTCCAATGCTGTCTTTGGCCAGA 2070
Qy 741 AspAlaAlaAsnLeuAlaValGlnGlySerGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2071 GATAGTGAACCTCACCAGAGCAGCAGAGGGGCTTTGGGCCAGCATCTTGAGAAT 2130
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2131 GAGACCTATGAGCAGCATCTCAGTGGCAGCGCCCTCTCCAGTCATACCCCCAGC 2190
Qy 781 ThrThrSerAlaProProLeuProProArgGlnValGlyLysValGlnThrAlaSerSer 800
Db 2191 ACCACGACGACCCCGCTCCACCTCGGAATGTTGGCAAA----- 2232
Qy 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2232 ----- 2232
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2233 -----GAT 2235
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2236 CCCCTGACCAACACACACCCCTCTCTGTGGTAAAGACATCAGGAACCTTAGAAGCCATG 2295
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2296 AACCAGCGGAGCAAGTCTCTCCAACTCGGGACCTCACAGAGCAGCCCGCCCTGCGCA 2355
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2356 CCTCAGCCACCCAGCGCGCTCGCTCAGAAAGAACCCAGCTTCCGGAGCTGACAGCCCAAC 2415
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 2416 CCGCTGACCAACAAAGGACGACGCGAGAGGACCT-----GAAGCTTCG 2457
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2458 GGTCTCTGTCCAAGCCATGGCTCGAGCCCGCCAGCAGCCATCGCTTAGGAAGTCACAG 2517
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 2518 GCGACCAAGTCAAGCCCAAGCGAGTGAAGCAGCTCTACAACTGGCTGGCTGACAAACCA 2577
Qy 961 AspGluLeuThrPheSerGluGlyAspValLysValLysValLysValGluAspGlnGlu 980
Db 2578 GACGAGCTGACTTTCTCTGAGGGCGACGTGATCATTTGGATGGGGAAGAGGATCAGGAG 2637
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 2638 TGGTGGATCGGCCACATCGATGGAGAGCCCAAGCGCCCAAGGAGCGTTTCCCGTATCGTTT 2697
Qy 1001 ValHisPheIleAlaAsp 1006
Db 2698 GTGCACITTAATGCTGAC 2715

```

```

RESULT 6
AR243296
LOCUS AR243296 2949 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 8 from patent US 6475778.
ACCESSION AR243296
VERSION AR243296.1 GI:27290454
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

```

```

REFERENCE 1 (bases 1 to 2949)
AUTHORS Roberts,T.M., King,F.J., Harris,D.F., Hu,E., Spiegelman,B. and Chan,J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A B 05-NOV-2002;
FEATURES
    source
        1. .2949
        /organism="unknown"
        /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 4.19e-136 Length: 2949
Score: 3874.00 Matches: 755
Percent Similarity: 83.12% Conservative: 92
Best Local Similarity: 74.09% Mismatches: 122
Query Match: 73.47% Indels: 50
DB: Gaps: 13
US-09-914-042-1 (1-1006) x AR243296 (1-2949)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 1 ATGCCTGACCAGATAACAGTGGCGGAGTTGTCCAGGAGACAAATGAAGATTATAAATCG 60
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 61 CCCACCCTCTCAAACTTCACCACCAGAAATGACTCACTGCAGAAACACAGTATCCGACTG 120
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 121 GAGGAGCCCTGGATGTGGACCGCAGTCTCTTTACAAGATGAAGAGTCAAGTAAAGCT 180
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 181 ATTTACGCTCGGCTCGGCTCATGTGAGAAATGAGGAGCAGTACACTCAAGCTCTGGAG 240
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 241 AAGTTCCGAGAGAACTGTGTGTACAGAGATGACCCGACCTGGGATCAGCTTCTGAG 300
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 301 TTCTCCGCTTTCACCAAGGAGCTCAGCGCACTCTTCAAGAACCTGTTTCAGAACATGAT 360
Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 361 AATATCATTAACCTTCCCATTTGGACAGTCTGCTGAAGGAGATCTGAAAGGGGTTAAAGG 420
Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrLysIle 160
Db 421 GATCTCAAGAGCCCTTCGATAAAGCCTGGAAGACTACGAGACTTAAGTCTCTAAATA 480
Qy 161 GluLysGluLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 481 GAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGlu 200
Db 541 GCAGAGATAGCAGAGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 201 TyrLeuLeuLysValAsnGluIleLysLysLysValAspLeuLeuGluAsnLeu 220
Db 601 TACCTCTCAAGTCAATGAATCAAGATCAAAAGAGGTGCGACCTGCTCCAGATCTC 660
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 661 ATCAATATCTTCCAGCAGCAGTGAACCTCTTTTCCAGGATGCTCTCAAGCGGTGGACAAC 720
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 721 CTCAAGCCCTCAATAGAAAAAAGCTGCCACAGACTTGCACTCGATCAACAGGTACAGGAT 780

```

Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 781 GAAGAACGACAGACAGTAAACCGATTACGGATGTGTAAAAATGCTCTGCAAGTGGAG 840  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 841 CAGAAAGGAGGACTCTCAGGTTAGACAGAGCGCCACCTACAGTCTGCACAGCGCGGCG 900  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 901 AACAAAGAGCATGGGACTGAGCGCAGCGCAACCTTTACAGAGAGTGAGCGGCTGGG 960  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 961 AAAGTGTGGCAGAGAGAGAGTGCACAGTAAGAAATGGATATTTGACCATCTCACATGGG 1020  
Qy 341 ThrAlaAsnArgProProAlaLysLeuHisLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1021 ACGGCAAAACACACCTCCCGCAAACTCAATCTTCTCACCTGTGAGGTGAAGCACAACCCA 1080  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1081 GAGGAGAGAGAAAGTTTGACCTCATCTCATGACAGAAATATCATTTCCAGGCAGAA 1140  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1141 GATGAGCAGAGTGTCAAAATATGATCTCAGTGTGTCAGACAGTAAAGAGAGCGGCTC 1200  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1201 AACACGCTCTTACAGGGCGACAGCATGTTGGTGAATAATCAATTTGTCGAGAGCTCAC 1260  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1261 AAGGCCATCTGGGAGAGGTGAGCGGATGCGGGGAACGATGCTGTGCGAGTCCGGT 1320  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1321 GCTCCGCGCCACATGCTCTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTCG 1380  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1381 GGGATCCACAGAGAGCTGGGCGCTCATTTACTCCGAATCCAGTCCCTCACACTCGACGTC 1440  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1441 CTCAGCACTCGAGCTCTTGCTGGCCAGAACGTTGGGGAATGCTGGCTTCAATGAGATC 1500  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1501 ATGGAGGCGCTGTCTGACGGCAGAGATGTGATCAAAACCGAATCCAGCCAGTGACATGCAG 1560  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1561 GCGAGAGAGGACTTTATCATGCCAAATACACAGAGAAACGCTTCGCTCGTAAGAAGTGT 1620  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1621 CCAGCGCAGCTGTGAAAGCTGCACAGCTGTGTGATGCTGTGAAGCCCGGACATTTTC 1680  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1681 TCTCTCATCCAGGCTATGCTGAAGAGTGGATCTGATGGAGCCCATTTCTCTGCTAAT 1740  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 1741 GGCATGAAACAGGTGAGACGCTCTTCATCTGGCCGTGAGACTGGTGAGACAGAACTTCC 1800  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 1801 CTACACATCATGAGCTTCTTACCCCAAAACAGTTTAAACCTGGATGAAGCAACCGCTAAA 1860  
Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640

Db 1861 GGAAGCAGAGCTGTCATTTACTGCTGCTGCGCAACACAGCGAGTGTCTCAAACTGCTG 1920  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660  
Db 1921 CTCAGAGAAAGAGCTTCATAGATATCGTANTAGCTGGAGACCCCGTTGGACATC 1980  
Qy 661 AlalysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 1981 GCCAGGCGACTCAAACTCTCAGTGTGAGAACTGCTGAAACCAAGGCTCTTGCAGGGAAG 2040  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2041 TTCAATGCTCATGTGATGTGAGTATGAGTGGAGACTTCAGCATGAAGACCTGACGAG 2100  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2101 AGTATGAGATCTGGATGAGATGCGAGTCTCT-----CACCGCGGAGTGGCGGCC 2154  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2155 ATCAGCTGTACACACCGGCGCAGTAACCTCTTCAGCTGAGTCCAGCGCAGCTGAGCGCA 2214  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2215 GACGCTCAGAGACCTGGTTAAAGACAGCAACGCG-----TTTGTGCCAAACCTGCTCAACAT 2271  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2272 GAAACCTACGGGAGCATCATTAACACACCGACTCACCC-----GTCAGCGCTG 2316  
Qy 781 ThrThrSerAlaProProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2317 TCCTCTTCTGCTCCACCTCTACCAACCCCGAAGAACCTA-----GTTCCGCGCTGCTCTT 2370  
Qy 801 AlaAsnThrLeu-----TrpLysThrAsnSerValSerValAspGlyGly 815  
Db 2371 GCAGAGCTGACTCAAGGATCTCCGCGCTGGAAGCGCTGCTCTCTGGATCTGAGCGCGC--- 2427  
Qy 816 SerArgGlnArgSerSerSerAspProAlaValHisProProLeuProProLeuArg 835  
Db 2428 ---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTCGCGCTCCCTTACGG 2484  
Qy 836 ValThrSerThrAsnProLeuThrProThr---ProProProProValAlaLysThrPro 854  
Db 2485 GTCACTTCCACTCCCTTCTAATGCCAGCGGTGCTGCTCTCTCTCTGCTAAAGCTACT 2544  
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyIleSer 872  
Db 2545 GGTATGATGGAGACCATGATATGCAACCCAAACCCGAGAGCGGCTCTCTGGACAGAAC 2604  
Qy 873 GlnIleArgProProProLeuProGlnProProSerArgLeuProGlnLysLysPro 892  
Db 2605 ATCAACCGG----- 2613  
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 908  
Db 2614 GCTACAGTGGGACAAAAAACTTCAGCAAAAGCACACTGATGCGCTCC----- 2661  
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927  
Db 2662 ---GGATCATCTGAGAGACCATGATATGCAAGAGTCCAGGAGGCGCCACAAAACACCACTGGT 2718  
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLysLysProLys 947  
Db 2719 CAAACTCTGCTGGACCCCATGCCCCAGGAAA-----ACGTATTGAGCGCGAAG 2769  
Qy 948 ArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGlu 967  
Db 2770 CGTGTGAAGGCATGTATTAATCTGTGGCCCATTAATCCAGACGAGCTGACCTTCTCTGAG 2829  
Qy 968 GlyAspValIleIleValAspGlyGluAspGlnGluThrTrpIleGlyHisIleAsp 987



Db	2830	GGAGAGCTTATCGTGGTGGATGGAGAGGAGGAGCACCAGGAGTGGTGGCTGGGCCACCATTTGAG	2889
Qy	988	GlyAspProGlyAArgLysGlyValAlaPheProValSerPheValHisPheIleAlaAsp	1006
Db	2890	GGAGAGCCATGACAGAGGAGGCGTTTCCTGTCAGGTTTGACAGTTCAATTATGAC	2946
RESULT 7			
AR243295			
LOCUS	AR243295	5954 bp	DNA
DEFINITION	Sequence 6 from patent US 6475778.		
ACCESSION	AR243295		
VERSION	AR243295.1	GI:27290453	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5954)		
AUTHORS	Roberts, T.M., King, F.J., Harris, D.F., Hu, E., Spiegelman, B. and Chan, J.		
TITLE	Differentiation enhancing factors and uses therefor		
JOURNAL	Patent: US 6475778-A 6 05-NOV-2002;		
FEATURES	Location/Qualifiers		
source	1..5954		
	/organism="unknown"		
ORIGIN	/mol_type="genomic DNA"		
Alignment Scores:			
Pred. No.:	8.69e-136	Length:	5954
Score:	3874.00	Matches:	755
Percent Similarity:	83.12%	Conservative:	92
Best Local Similarity:	74.09%	Mismatches:	122
Query Match:	73.47%	Indels:	50
DB:	6	Gaps:	13
US-09-914-042-1 (1-1006) x AR243295 (1-5954)			
Qy	1	MetProAspGlnIleSerValSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla	20
Db	433	ATGCCTGACCAAGATAACAGTGGCGGAGTTTGTTCAGGAGACAATGAAGATTATTAATCG	492
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla	40
Db	493	CCCCCGCCTCAAACTTCACCACCAAGATGACTCACTGCAGGAACACAGTATCCGCACTG	552
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	553	GAGAGAGCCCTGGAGTGTGGACCGCAGTGTCTCTTTACAGAGTGAAGAAGTCAGTTAAGCT	612
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu	80
Db	613	ATTACGCCTCGGTCTGGCTCATGTGGAGAAATGAGGAGCAGTACACTCAAGCTCTGGAG	672
Qy	81	LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys	100
Db	673	AAGTTCGAGAGAACTGTGTGTACAGAGATGCCCGGACCTGGGATCAGCCTTCTCTGAAG	732
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	733	TTCTCCGCTTTCACCAAGAGCTCACGGCACTCTTTCAGAACCTGTCTTTCAGAACATGAAT	792
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	793	AATATCATTTACCTCCCAATTGGACAGTCTGCTGAAGGGGAGATCTGAAGGGGGTTAAAGGG	852
Qy	141	AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle	160
Db	853	GATCTCAAGAGCCCTTCGATAAAGCCTGGAAAGACTACGAGACTAAGTCTCTTAAATA	912
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	913	GAGAAGGAGAAAGAACGACGCGCCGCGAGCAGCGAATGATCCGAGCGAGATCAGCGGA	972



Db 2053 CGACGCGACTGCGAAGCTGCACAGCTCTGTGATGCTGTAAGCCCGGACATTTTC 2112  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2113 TCTCTCATCCAGCTCTATGCTGAAGAGTGGATCTGATGGACCCATCTCTCTGCTAAT 2172  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2173 GCACATGAACAAGGTGAGACGCTCTTCATCTGGCGGTGAGACTGGTGACAGAACTTCC 2232  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2233 CTACACATCATCGACTTCTCACCACAAACAGTTTAAACCTGGATGAAGCAACCGCTAAA 2292  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2293 GGAACACAGCTCTGATCTGCTGCTGACGACACAGCGAGTGTCTCAATGCTG 2352  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2353 CTCAGAGAAAGCCTCCATGATATCGTAAATGAAGCTGGAGACACCCCGTTGGACATC 2412  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2413 GCCAGCGACTCAACATCTGCAGTGTGAGGAACCTGCTGAACACAGGCTCTTGCAGGGA 2472  
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2473 TTCAATGCTCATGTGATGATGAGTATGAGTGGAGACTTCAGCATGAAGACCTGGACGAG 2532  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720  
Db 2533 AGTGATGAAGATCTGATGAGAAGTCGAGTCTCT-----CACCGCGGGATGAGCGGCC 2586  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2587 ATCAGCTGCTACACACCGCGAGTAACCTCTTCAGCTGAGTCCAGCGACGCTGAGCGGA 2646  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2647 GACGTCGAGACCTGGTAAAGCAAGCAACGC-----TTGTGCCAAACCTGGTCAACAA 2703  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2704 GAAACCTACGGGACCATTAACACACAGCTCACCC-----GTCAGCCTG 2748  
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2749 TCCTCTTCTGCTCCACCTTACACCCCGAAACCTA-----GTTCCAGCCGCTGCTCTT 2802  
Qy 801 AlaAsnThrLeu-----TrpLysThrAsnSerValSerValAspGlyGly 815  
Db 2803 GCAGAGCTGACTCAAGAGATCTCCCGGCTGGAAGCCTGGCTCTCTGGATCTGAGCGGC--- 2859  
Qy 816 SerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArg 835  
Db 2860 ---AGACAGAGATCTCTCTGACCTCCCAACATGCATCTCTCTCGCGCTCCCTTAGCG 2916  
Qy 836 ValThrSerThrAsnProLeuThrProThr-----ProProProProAlaLysThrPro 854  
Db 2917 GTCACCTCCACCTCTTAATGCCAGCGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2976  
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyIleSer 872  
Db 2977 GGTATGATGGAGACCATGAATATGCAACCAACCCGACAGCGGCGCTCTCTGACAGAAC 3036  
Qy 873 GlnIleArgProProProLeuProGlnProProSerArgLeuProGlnLysLysPro 892  
Db 3037 ATCAACCGG-----LysSerThrProLeuThrAsnLysGlyGlnPro 908  
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 942

Db 3046 GCTACAAAGTGGGACAAAACACTTCAGCAAAAGCACACTGATGCGCTCC----- 3093  
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927  
Db 3094 ---GGATCATCGAGAGACGACTAAAGAGTCCCGAGGAGGCCACAAAACACCACTGGT 3150  
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947  
Db 3151 CAAACTCTGCTCGCAGCCACATGCTCCAGGAAA-----ACGTATTGAAGCGGAAG 3201  
Qy 948 ArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGlu 967  
Db 3202 CGTGTGAAGGCCCATGATAACTGTGTGGCCGATATCCAGACGACTGACCTTCTCTGAG 3261  
Qy 968 GlyAspValIleIleValAspGlyGluAspGlnGluTyrTrpIleGlyHisIleAsp 987  
Db 3262 GGAGAGCTTATCGTGGTGGATGGAGAGGAGGACAGGAGTGGTGGCCACATTGAG 3321  
Qy 988 GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006  
Db 3322 GGAGAGCCAATGAGAAGAGGAGCGTTCTCTGTACAGTTGTACAGTTTCATTATGGAC 3378  
RESULT 8  
LOCUS I74317 4486 bp DNA linear PAT 03-APR-1998  
DEFINITION Sequence 15 from patent US 5688641.  
ACCESSION I74317  
VERSION I74317.1 GI:3010458  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4486)  
AUTHORS Sager, R., Zou, Z., Lee, S. Whan, and Tomasetto, C. Laure.  
TITLE Cancer diagnosis using nucleic acid hybridization  
JOURNAL Patent: US 5688641-A 15 NOV-1997;  
FEATURES Location/Qualifiers  
source 1..4486  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,35e-129 Length: 4486  
Score: 3691.00 Matches: 701  
Percent Similarity: 99.29% Conservative: 1  
Best Local Similarity: 99.15% Mismatches: 3  
Query Match: 70.00% Indels: 2  
DB: 6 Gaps: 0  
US-09-914-042-1 (1-1006) x I74317 (1-4486)  
Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLys 321  
Db 1 AAGGAACATGGAGCCGAGCGGAACCGGAGCTCTACAAGAAGAGTGACGGGATCCGAAA 60  
Qy 322 ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341  
Db 61 GTGTGGCAGAAAAGAAATGTTTCAGTTAAATGGTTTTCTGACCATATCCCTGGTACC 120  
Qy 342 AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGly 361  
Db 121 GCTAACCGGGCTCTCTGCAAGAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCTGA 180  
Qy 361 uGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAs 381  
Db 181 GGAGAAGAAGTGTCTTGGACCTTATTTCATCATGACAGAACTTACCACCTTTCAGCTGAAGA 240  
Qy 381 pGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeuAs 401  
Db 241 TGAACAGGAATGTCAATATGATGCTGTGCTGCAAAATAGCAAGAGAAAGAGCTTTAAA 300  
Qy 401 nAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrIly 421

```
Db 301 CAATGCTTTAAGGGGATGACAATACTTGGAGAAATAAATGCTCCAAAGAACTGACAAA 360
Qy 421 sGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAl 441
Db 361 GGAGATCATCTCAGAAGTGCAGAGGATGACGGCAATGACGCTGCTGTGACTGTGGGGC 420
Qy 441 aProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysValIleGluCysSerG1 461
Db 421 GCCAGATCCTACATGGCTTTTCACCACTCTGGGCATCTCTGACCTCGAGTGTTCGG 480
Qy 461 yIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 481
Db 481 AATCACCGAGAGCTGGGGTTCATTATTCCAGGATGCAGTCCCTGAGCTTAGATGACT 540
Qy 481 uGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnIleGlyAsnIleGlyPheAsnGluIleMe 501
Db 541 GGGAAACATCTGAGCTGCTCGCCCAAGAAATATTGGGAATGCAGGCTTTAATGAGATCAT 600
Qy 501 tGluCysCysLeuProAlaGluAspSerValIysProAsnProGlySerAspMetAsnAl 521
Db 601 GGAATGTTGCCCTACAGCTGAGGACTCAGTCAACCCCAACCCAGGACGACATGAATGC 660
Qy 521 aArgIysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAl 541
Db 661 AAGAAGAGCTACATCACAGCAAGTACATCGAGAGGAGATACGCAAGAGAGACGACGC 720
Qy 541 aAspAsnAlaLysLeuHisSerLeuCysGluAlaValIysThrArgAspIlePheG1 561
Db 721 GGATAACCGCGCGAAGCTTCAGTCTTTGGAGGCGCTCAAAACGAGAGATATTTTTGG 780
Qy 561 yLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluIysIleProLeuAlaAsnG1 581
Db 781 ATTGCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAATCCCACTGGCCACGCG 840
Qy 581 yHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLe 601
Db 841 ACATGAGCGGATGAAAGCGCCCTCACCTTGCAGTCAGATCCGTGGATCGAACCTCTCT 900
Qy 601 uHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspIysGlnThrGlyLysG1 621
Db 901 TCACATTGTAGACTTTTGTAGTTTCAAGACAGTGGGAACCTGGATAAACAGACAGGGAAAG 960
Qy 621 ySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 641
Db 961 CAGCACGCCCTGCACTACTGCTGCTGACCGACAATGCCAGTGCCTCAAGTTGCTCCT 1020
Qy 641 uArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAl 661
Db 1021 CGGGGGAGGCTCCATCGAGATAGCAATGAGTCAGAGAGACTCCGCTGGACATTCG 1080
Qy 661 aLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPh 681
Db 1081 CAAGCGCTCAAGCAGCAGCACTGTGAGGAGCTGTGACCCCAAGCTTATCTGGAAGATT 1140
Qy 681 eAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGluSe 701
Db 1141 TAAATCTCACGTTTCACTGTTGATATGAATGGCGACTACTCCACGAGACCTTGGATGAAAG 1200
Qy 701 rAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgProI1 721
Db 1201 TGATACGACATGGATGAGAAATTCAGCCAGTCCCAACCGCGGGAGACCGGCCCAT 1260
Qy 721 eSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAs 741
Db 1261 CAGCTTCTACCACTGGGCTCCCAACGAGCTTCAGTCTAACGCTGATCTTTGGCCAGAGA 1320
Qy 741 pAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnG1 761
Db 1321 TGCTGCAAACTTGGCAAGGACCAAGCAGAGGCTTTTCATGCCCAGCATCTTGCAGAAATGA 1380
Qy 761 uThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSerTh 781
```

```
Db 1381 GACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCCGCCAGCCTCGACCCCCCAGCAC 1440
Qy 781 rThrSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSerA 801
Db 1441 CACAGCGCCCCCCCCCTCTCCACGAATGTGGCAAGTTTCAGACACGCTCCTCTG 1500
Qy 801 laAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSers 821
Db 1501 CTAACACCTGTGAAGACAACTCTGTAAGTGTGGAGCGGTGGAAGCGGCGAGCATCTT 1560
Qy 821 erSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsnP 841
Db 1561 CGTCAGATCGCCAGCTGTCCATCCACCGCTGCCCTCTTCGCGTGCACATCTACCAATC 1620
Qy 841 roLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeuS 861
Db 1621 CCCTGAGCCCCACGCCGCCCTCCCGTGCAGAGCGCCGCGTGAATGAAGCCTTGA 1680
Qy 861 erGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuProP 881
Db 1681 GCCAGCGCAGCAAGCTTCCCCGCTGGATCTTCAGAGATCAGGCCCCCCCCTCTGCCCC 1740
Qy 881 roGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrp 901
Db 1741 CACAGCGCCCGCAGCGCTCCCGCAGAAGAAGCCTCGCCAGGGGCTGACAAGTCCACCC 1800
Qy 901 roLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuG 921
Db 1801 CACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCTGG 1860
Qy 921 lyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGlnA 941
Db 1861 GTCTCTGTCCATGCTATGCTCTGACGCCCCCTGACCCATGCTAGGAAGTCGCAGG 1920
Qy 941 laThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProA 961
Db 1921 CAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGCTGACAACCCCG 1980
Qy 961 spGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGluT 981
Db 1981 ATGAGCTCACCTTCTCCAGGGGATGTGATCATCTGTGAGCGGGAGGAGACCGAGGT 2040
Qy 981 rpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheV 1001
Db 2041 GGTGGATTGGCCACATGTATGGAGATCTCTGTGTCGCAAGAGCGCATTCGCGTGTCAATTG 2100
Qy 1001 alHisPheIleAlaAsp 1006
Db 2101 TGCACCTTTATCGCTGAC 2117

RESULT 9
I74314 4328 bp DNA linear PAT 03-APR-1998
LOCUS Sequence 12 from patent US 5688641.
DEFINITION I74314
ACCESSION I74314
VERSION I74314.1 GI:3010455
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 4328)
TITLES Seger R., Zou, Z., Lee, S. Whan, and Tomasetto, C. Laure.
JOURNAL Cancer diagnosis using nucleic acid hybridization
FEATURES Patent: US 5688641-A 12 18-NOV-1997;
Location/Qualifiers
1..4328
source /organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores: 6.98e-119 Length: 4328
Pred. No.: 3417.00 Matches: 648
Score:
```

Percent Similarity: 99.39% Conservative: 1  
 Best Local Similarity: 99.23% Mismatches: 3  
 Query Match: 64.80% Indels: 1  
 DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x I74314 (1-4328)

QY 355 GlnValLysThrAsnProGluLysLysCysPheAspLeuLeuSerHisAspArgThr 374  
 DB 3 CAGGTGAAGACCAACCTTGAGAGAGAGAGTCTTTGACCTTATTTACATGACAGAACT 62  
 QY 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTropMetSerValLeuGlnAsn 394  
 DB 63 TACCACTTTCAAGCTGAGAGTGAACAGGAATGTCAAAATATGATGTCTGTGTCGCAAAAT 122  
 QY 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsn 414  
 DB 123 AGCAAGAAGAGCTTTAAACAATGCATTTAAGGGGGATGACAACTTGGAGAAATTAAC 182  
 QY 415 IleValGlnGluLeuThrLysGluIleLeuSerGluValGlnArgMetThrGlyAsnAsp 434  
 DB 183 ATCGTCCAGAACTGACAAAGAGATCATCTCAGAAGTGCAGAGATGACGGGCAATGAC 242  
 QY 435 ValCysCysAspCysGlyAlaProAspProThrTropLeuSerThrAsnLeuGlyIleLeu 454  
 DB 243 GTCTCTGTGACTGTGGGGCGCCAGATCTCATGGCTTCCACCAACTGGGCATCTCTG 302  
 QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474  
 DB 303 ACCTGCATCGAGTGTCCGGAAATCACCAGAGAGCTGGGGTTCATTTCCAGAGATGAC 362  
 QY 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsn 494  
 DB 363 TCCCTGACCTTAGATGTCTGGGAACATCTGAGCTGTCTCGCCCAAGAAATATGGGAAT 422  
 QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsn 514  
 DB 423 CGAGCTTTAATGAGATCATGGAATGTTCCTTACCAGCTGAGGACTCAGTCAAAACCCAAC 482  
 QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArg 534  
 DB 483 CCAGCAGGACATGATGATGCAAGAAGAGCTACATCAGCCCAAGTATCATCGAGAGGAGA 542  
 QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaVal 554  
 DB 543 TACGCAAGGAGAGAACACCGGATTAACCGGGGAAGCTTCAAGTCTTTTGGAGGCCGTC 602  
 QY 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574  
 DB 603 AAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGGTGTGGATCTTACGGAA 662  
 QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594  
 DB 663 AAAATCCCACTGCGCCACGACATGAGCGCGGATGAACGGCCCTCCACTTGCAGTCAGA 722  
 QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614  
 DB 723 TCCGTGGATCGAACCTCTCTTCCACATTTAGACTTTTATGTTTCAAGACAGTGGGAACCTG 782  
 QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634  
 DB 783 GATAAACAGACAGAGGAAAGGACAGCAGCCCTGCACTACTGCTGCTTGCACCAATGCC 842  
 QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654  
 DB 843 GAGTCCCTCAAGTCTCTCTCGGGGGAGGGCTCCATCGAGATAGCAAAATGAGTCAGA 902  
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674  
 DB 903 GAGACTCCGCTGACATTTGCCAAGCGCTCAGACAGAGCACTGTGAGGAGCTGTGACC 962  
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluThrPArgLeuLeu 694

DB 963 CAAGCCTTATCTGGGAAGATTTAAATTTCTCAGCTTCACTGTTGAATATGAATGGCAGCTACTC 1022  
 QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714  
 DB 1023 CACGAAGACCTGGATGAAGTGTATGACGACATGATGAGAAATTCAGCCCACTGCCAAC 1082  
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734  
 DB 1083 CGCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCCAACAGCTTCACTCAAC 1142  
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754  
 DB 1143 GCTGTATCTTTGGCCGAGATGCTGCAAACTTGGCAAGACACAGAGGGCTTTTCATG 1202  
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774  
 DB 1203 CCCAGCATCTTGCAGAAATGAGACTTTACGAGGCCCTCTCTGAGTGGCAGCCCACTCCCGCC 1262  
 QY 775 GlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794  
 DB 1263 CAGCCTGAGCCCGCCAGCACCACAGCCGCCCTCCCTGAGTGGCAGCCCACTCCCGCC 1322  
 QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuTropLysThrAsnSerValSerValAspG 814  
 DB 1323 AGTTGAGACAGCTCTCTCTGTCTAAACCTGTGGAAGACAACTCTGTAAGTGTGACCG 1382  
 QY 814 yGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLe 834  
 DB 1383 TGGAAAGCCCGCAGCGATCTTCTGTCAGATCCGCCAGCTGTCCATCCACCGTCCCTCT 1442  
 QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPr 854  
 DB 1443 TCGCTGACATCTTACCAATCCCTGACCCCAAGCCGCCCTCCCTGTTGCCAAGACGCC 1502  
 QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnI 874  
 DB 1503 CAGCGTAATGAAGAGCTTTGAGCCAGCCAGCAAGCTCTCCCGCTTGGATCTCACAGAT 1562  
 QY 874 eArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894  
 DB 1563 CAGGCCCCCACTCTGCCCCCAGCCGCCCTCCCGCAGAGAGAGAGCTTGGGCC 1622  
 QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLe 914  
 DB 1623 AGGGCTGACAAAGTCCACCCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGATCT 1682  
 QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934  
 DB 1683 CTCTGCAACGGAAGCTCTGGGTCTCTGTCATGCTATGGTCTTGCAGCCCTTGCACC 1742  
 QY 934 oMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAs 954  
 DB 1743 CATGCTAGGAAGTCCGACGGCAACCAAGTTGAGCCTAAGCGGTGAAGAGCTCTATAA 1802  
 QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAs 974  
 DB 1803 CTGTGTGCTGACAAACCCCGATGAGTCACTTCTCCGAGGGGATGTGATCATCGTGA 1862  
 QY 974 pGlyGluGluAspGlnGluTropIleGlyHisIleAspGlyAspProGlyArgLysG 994  
 DB 1863 CGGGAGAGAGACAGAGAGTGTGGTGGCCACATGATGAGATCTCTGTGCGCAAGG 1922  
 QY 994 yAlaPheProValSerPheValHisPheIleAlaAsp 1006  
 DB 1923 CGCATTCGGGTGTCATTTGTGCACTTTATCGCTGAC 1959

RESULT 10  
 AR243298  
 LOCUS  
 DEFINITION Sequence 11 from patent US 6475778.  
 ACCESSION AR243298  
 VERSION AR243298.1 GI:27290456  
 KEYWORDS

AR243298 2712 bp DNA linear PAT 20-DBC-2002



```
Db 1798 CTCCTATTGTGGACTTCTTGTACAGAACAGTGGCAATTTAGACAGACAGACCCAAA 1857
QY GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 1858 GGAAGCACAGCGCTACATATTGCTTGTGCTGATGATACAGTGAATGATGAAGTGTCTG 1917
QY LeuArgGlyLysAlaSerIleGluLeuAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 1918 CTGCGGGGGAAGCATCTGTGACGATTAATGATGCTGGAGAGACTGCTCTGGATTG 1977
QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 1978 GCGCAGCGTCTCAACACTCCAAATGCGAGGAGCTGCTGACTCAGCGCAGACGGGGAG 2037
QY 681 PheAsnSerHisValHisValGluTyrGluThrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2038 TTCAATGTCATGTGATGTAATGACTGGCGTCTGCATAATGAGGATCTGGACGAG 2097
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2098 AGCGAAGATGAGATGGAGACAAG-----CCATTCCCATCAGGCGTGAGGAGGTCCA 2151
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2152 ATAACTGTATAGTTCAGGCACTGGCCCATGATGCCCAACATCAGGCGCTCTGGCTCG 2211
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2212 GACGTGGCCAAATGTGTCATTAATAGCAGAGGGCTTTATTCCGAGCATGATGATGAC 2271
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780
Db 2272 GAGACTTACGGCACCATGCTCGATCCCAACTCTCCACCACCTG-----GGTTTACCAGGA 2325
QY 781 ThrThrSerAlaProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2326 GTACCTGGCATTCCTCTTTACCCCTCGGCCCTTGGGAAGGGGA----- 2370
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2371 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGGTCA 2409
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2410 TGTTCAGATCTCTGCA-----AACCTCTCAAACTCTCTGAACAAAATAAATCTCTGTAT 2460
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2461 GTTCTGCTCTGCTCTCTCCACTCTCTCTGCA----- 2493
QY 861 SerGlnProSerLysProAlaProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2494 -----CCCAAGAGACTCCACTCCAGATCCAAAGGCCAGTCTTCTCTCCAGCAGCC 2547
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2548 ACGGCTCTCT-----CCTGCACCA-----TCCGCA 2571
QY 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 2572 CGCGCTCTTATTCACCTGCTCTCTCAGCGCAGCGCT----- 2610
QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2611 -----GTAGTGCCCTCTGCACCAAGTTATGCCCACTTCGTC 2646
QY 941 AlaThrLysLeuLysProLysArgValLysAla 951
Db 2647 CTGACTGATGTCAAAGTCTGCTGCTAAAGCC 2679
RESULT 11
AR243297 4595 bp DNA linear PAT 20-DEC-2002
LOCUS
```

```
DEFINITION Sequence 9 from patent US 6475778.
ACCESSION AR243297
VERSION AR243297.1 GI:27290455
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4595)
AUTHORS Roberts,T.M., King,F.J., Harris,D.F., Hu,E., Spiegelman,B. and Chan,J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A 9 05-NOV-2002;
FEATURES Location/Qualifiers
source 1..4595
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 4.41e-109 Length: 4595
Score: 3155.00 Matches: 621
Percent Similarity: 75.60% Conservative: 98
Best Local Similarity: 65.30% Mismatches: 174
Query Match: 59.83% Indels: 58
DB: Gaps: 10
US-09-914-042-1 (1-1006) x AR243297 (1-4595)
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 300 ATGCTGTATGCTGACGTCAGCTGAGTTGTGGATATATTACCCATGAGGACTATAAAGCA 359
QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 360 CCGCAACATCAGTCTTCTGCACGCGCATGCTCACTGCAGGAATACAGTCCCGCTCTG 419
QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 420 GAAGAGGCGCTGGATCTCGACCGCAGTGTACTGCACAAAATGAAGAAGTCACTCAAGGC 479
QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 480 ATAAACAGCTCTGGTCAGACTCATGTAGAACCGAGGAGCAGTACATCCAGGCCATAGAG 539
QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 540 AGGTTACGGATAACACTGTGTACAAAGTACCTCTGAGATGTCCAAATTAATCTCTCACA 599
QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 600 TTCGCTGTTTCCACCAAGGAGCTTACTGCTCTTTTCAAGAACTTCTGTACAGAACATGAAT 659
QY 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 660 AACATCATCACTTTTCCACTAGCAGTCTGTAAAGGGAGACCTCAAGAGGAGTCAAGGG 719
QY 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 720 GATTGAAAAGACCATTTTGCATTAAGCATGGAGGATTAATGAACCAACTGAGCAAGATT 779
QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 780 GAGAAAGAAAAGCGAGAACATGCCCAACAGCACGCTCTGTATCCGAACAGAGATCATGTCGA 839
QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 840 GGAGAGATCGCAGAGAGATGGAGAAAGAGAGACGCTGTTTTCAGCTTCAGATGTGTGAG 899
QY 201 TyrLeuLeuLysValAsnGluIleLysIleLysValLysGlyValAspLeuLeuGlnAsnLeu 220
Db 900 TACCTCATTAAGTGAATGAATCAAGTCAAAAGGGGGTGCAGCTCTCTTCACACACCTC 959
QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
```

Db 960 ATCAAACTACTTTTCATGCCAGTGCATTTCTTTCAGGATGGCTAAAGTCTGTGACCAAT 1019  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1020 CTGAAACCTTTTCATGAAAGCTTTCACAGACTTAACCGCGAAACAAACAGACTCAAGAT 1079  
Qy 261 GluGluArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1080 GCAGAAAGCAACAGTCTGTGAGCTGAAAGAACTCTTAAATCTGTCTACAGTCTGAG 1139  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 1140 TGTAAGGAGGATGCTCAGTCAAGCAGAACGACGAGCTACAGTCTTCCACGAGTTCGAGGC 1199  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArg 320  
Db 1200 AATAAAGCTCAGGCAGCGAGCGCTCTGGGATGCTCTCAACGCGAGCGAGACTGAGG 1259  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1260 AAAGTTTGGCAGAAAGGAAGTCTCTGTGAAATGGATGTTGACTATTTCACATGGA 1319  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360  
Db 1320 ACGCCCATGCACCGCCAGCAACCTGAACCTTTAACTGCGCAAGTGAAGGTAAACCA 1379  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1380 GATGAGAAAAATGCTTTGATCTCATATCATATGATGACAGAACGATATCACTTCCAGACTGAG 1439  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1440 GATGAGGAGAGTGTGAGTATGGGTCTCTGTCTCCAGAACAGTAAAGAGGCGCTG 1499  
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1500 AACATGCTTTAAAGACGATCAGAAATGAGGGAGAGAAATAACATGTTTCGAGAGTCACT 1559  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1560 AAGGCATCTGTGGGGAAAGTGAAGAAATGAGCGCAATGAGCTGTGCTGACTGTGGA 1619  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1620 GCTTCCAATCCAAACATGGCTCTCCAAACCTGGGTGTGTTGATTGCAATGAATGCTCT 1679  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1680 GGGATCCATCGGAAATGGCGTCCACTACTCCCGAATACAGTCTTGACACTGGACCTC 1739  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1740 TTAGGCACATCTGAACATTATTGCTTAACAGTGTGGGAAATGACAGCATTTCAATGAAATC 1799  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1800 ATGGAAGCAAACTGTTCTCAGAG---ATCCCAAAACCCCTACCTTCTAGTGACATGAG 1856  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1857 GTACCAAAAGACTTCATCACACCCAAATACACAGAGAGCGTTTCGCTCAGAGAAGTAT 1916  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1917 GCAGATAACGCACTCGACTGCATGCACTGTGTGATGACGTGAAGTGAAGTTCGCGACATCTTC 1976  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1977 TCCCTGATCCAGGCTTATGCTGAAGACTGGACCTGTGAGGACCATTAATCAGGCTAAC 2036  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600

Db 2037 CAACATGAACAGCGGAGACATCACTACATCTTGGGTACGAATGGTGGACCGAAACTCC 2096  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGluThrClyLys 620  
Db 2097 CTCCATATTGGGACTTTCTTGTACAGAACAGTGGCAATTTAGACAACGACACAGCCAAA 2156  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2157 GGAAGCAGCGCTACATTATTGCTGTGCTGATGATACAGTGAATGTATGAAGTGTGTG 2216  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2217 CTGGGGGGAAGCATCTGTGAGCATTAATGATGTCGAGAGACTGCTCTGGATTGT 2276  
Qy 661 AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2277 GCGCAGCGTCTCAAACTCCAAATGCGAGAGCTGTGACTCAGGGCCGACAGCGGGAAG 2336  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2337 TTCAATGTCCATGTGCATGTGGAATATGACTGGCGTCTGCATATATGAGGATCTGGACGAG 2396  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2397 AGCGAAGATGATGAGAGGACAAG-----CCCATTTCCATCAGGCGGTGAGGAGCTCCA 2450  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2451 ATNAGCTGTATAGTTCAGGCGAGTGGCCCCATGATGCCAACATGAGCGCTCTGGCTCGG 2510  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2511 GACGTGGCCATGTGGTCAATAATANGCAGAGGCGTTTATTCGAGCATGATGATGAAC 2570  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2571 GAGACTTACGGCAGCATGCTCGATCCCAACTCTCCACCACCTG-----GGTTTACCAGGA 2624  
Qy 781 ThrThrSerAlaProProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2625 GTACTGGCATCTCTTTTACCCTCGGCGCTTGGGAAGGGGA----- 2669  
Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
Db 2670 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGGTCA 2708  
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840  
Db 2709 TGTTCAGATCTCGCA-----AACCTCAAACTCTCGAACAAATAAATACTCTGTGTAT 2759  
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
Db 2760 GTTCTGCTCTGCTCTCTCCACCTCCTCTCGCA----- 2792  
Qy 861 SerGlnProSerLysProAlaProGlyIleSerGlnIleArgProProProProLeuPro 880  
Db 2793 -----CCCAAGAGACCTCCACCTCCAGATCCAAAGGCCAGTCTTCTCTCCACAGCC 2846  
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900  
Db 2847 ACGGCTCCT-----CCTGCACCA-----TCCGCA 2870  
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920  
Db 2871 CCGTCTCTTATTCACACCTGCTCTCTCAGGCGAGCGCT----- 2909  
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940  
Db 2910 -----GTAGTGGCCCTCGACCATGATTATGCCCATCTTCGTCAC 2945  
Qy 941 AlaThrLysLeuLysProLysArgValLysAla 951  
Db 2946 CTGACTGATGTCAAAGTCTGCTGTCTAAAGCC 2978

```

RESULT 12
BC048818
LOCUS
DEFINITION
Mus musculus development and differentiation enhancing, mRNA (cdna
clone IMAGE:3668778), partial cds.
ACCESSION
BC048818
VERSION
BC048818.1 GI:28981428
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4434)
Strausberg, R.I., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, K.H., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Rulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bonfard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smallos, D.E.,
Schneringer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4434)
Strausberg, R.
Direct Submission
Submitted (14-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc\_mgc@nigri.nih.gov
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karins, E., Kwong, P., Laric, R., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 99 Row: a Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1. .4434
/organism="Mus musculus"
/mol_type="mRNA"
FEATURES
source
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3668778"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP_Mam5"
/lab_host="DHI08"
/notes="vector: pCMV-SPORT6"
<1. .4434
/gene="Ddef1"
/notes="synonyms: PAP, a19"
/db_xref="LocusID:13196"
/db_xref="MGI:1342335"
<1. .3504
/gene="Ddef1"
/codon_start=1
/product="Ddef1 protein"
/protein_id="AAH48818.1"
/db_xref="GI:28981429"
/db_xref="LocusID:13196"
/db_xref="MGI:1342335"
/translation="APLAPLLASPLARQAGRRGGCGARGVHEPPGPGSGRRRDRSRP
PPLVSRPSRGMPDQISVSEFIATETEDYNSPTSSFTRLHNCENTVTLLEALDQD
RTALQVYKSKVAINSGQDQVHVEENYAQVLDKFGSNFLSRDNPDLGTATFKFKLEKE
KREHAKGHMIRTEITGAEIAEEMKERRFLQOMCEYLIKVNETIKTKGVDLQNLQI
KYHAQCNFFQDGLTKADLKQYIEKLAADLYNIKQDEBKQQLTALRDLIKSSQL
DKTSOSRQGGYSMHQOGKEYGSKGFLKKSGDGRKVRWRRKCAVKNIGLITIS
HATSNRPAKLLNLTQCKVNAEDKKSFDLISHNRTVHFOAEDQDYATMISLVLTNSK
EALTMAFRGQSQTGENSLEDTKAIEDVORLPONDICCGSGSEPTWLTNLTGILT
CIECSGHRMGVHISRIELEDKLGTSSELLAKNVGNNSFNDIMEANLPSPPSKPT
PSSMTVRKEYITAKYVDHFRSRTKASSAKJNELLEAIKSRDLALIQVYAEVGL
MEPLLEQELGETALHVAQTQSLHLVDVFNQCNLDKQTSVGNTVLHVSGMY
GKPECLKLLRSKPTVDIVNONGETALDIKRLKATOCEDLLSQAOKSKFNPHVVEY
EWLRODEMDESDDLDKPSPIKERSPRQSSCHSSISPODKLALPGSTPRDKQ
RLSGATFNQIFASTSTDLPSFSEAPPLPRNKGKGTGPTPLGTQTSQSSST
LSKRRPPPPPPGPKHRTSLDPPSPPLPHGPNKGAIPWGNVGLPSSSKTANKEPGLSQQ
ASTSSAKTALGRVLPKLQPKVALRKTETSHLLSDRTNIPETFKSSQSLTELPPK
PLGLPLPKPVELAPKQVGLPELPPKGLDLPKQLDLPKQSLDLPKPKQMKDLPPK
POLGDLAKSOGDVSAKVQPPSVTQSHGTDLSPNVQSRDAIQKQASEDSNDLTPT
LPEPTPLPKINTGKNKVRVKTIYDCQANDDELTFIEGEVIIVTGEDEQEWIGH
TEGQPERKGVFPVSVFVHLSD"
1084. .1359
/gene="Ddef1"
/notes="PH; Region: Pleckstrin homology domain. Domain
commonly found in eukaryotic signalling proteins. The
domain family possesses multiple functions including the
abilities to bind inositol phosphates, and various
proteins. PH domains have been found to possess inserted
domains (such as in PLC gamma, sytrophins) and to be
inserted within other domains. Mutations in Brutons
tyrosine kinase (Btk) within its PH domain cause X-linked
agammaglobulinaemia (XLA) in patients. Point mutations
cluster into the positively charged end of the molecule
around the predicted binding site for phosphatidylinositol
lipids"
/db_xref="CDD:smart00233"
1420. .1761
/gene="Ddef1"
/notes="ArfGAP; Region: Putative GTPase activating protein
for Arf. Putative zinc fingers with GTPase activating
protein (GAPs) towards the small GTPase, Arf. The GAP of
ARD1 stimulates GTPase hydrolysis for ARD1 but not ARFp"
/db_xref="CDD:pfam01412"
1912. .2172
/gene="Ddef1"
/notes="ANK; Region: ankyrin repeats"
/db_xref="CDD:cd00204"
3325. .3492
/gene="Ddef1"
/notes="SH3; Region: Src homology 3 domains"
/db_xref="CDD:cd00174"

```



## Alignment Scores:

Pred. No.: 5,48e-104 Length: 4434  
 Score: 3018.00 Matches: 621  
 Percent Similarity: 65.88% Conservative: 128  
 Best Local Similarity: 54.62% Mismatches: 232  
 Query Match: 57.23% Indels: 156  
 DB: 10 Gaps: 18

US-09-914-042-1 (1-1006) x BC048818 (1-4434)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
 Db 166 ATGCGGACCAAGATCTCGGTGTCAGATTATCGCGGAGACACCGAGGACTACAACCTCG 225  
 Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValalaIle 40  
 Db 226 CCCACCACGTCAGCTTCACACGCGGCTGCACAACTGCAGAAACACCGTCACTGCTG 285  
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
 Db 286 GAGGGGCTCTAGACCAAGATAGACAGCCCTACAGAAAGTGAAGAAATCTGTAAGACA 345  
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyrThrGlnAlaLeuGlu 80  
 Db 346 ATATATAATTCGGGTCAAGATCATGTACAAATGAAGAAACTATGCAACAAGTTCTGGAT 405  
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
 Db 406 AAGTTTGAAGTAATTTTTTAAAGCAGACACACCTGACCTTGGCAGCGCTTTTGTCAAG 465  
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 Db 466 TTTTCTACATCACAAGGAAGTCTCCAGCTGCTGAAATATCTGCTCCAGGCTCTGAGC 525  
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140  
 Db 526 CACAATGTGATCTTCACCTTGATCTCTTTTAAAGGAGAGACTTGAAGGGGTCAAGGA 585  
 Qy 141 AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrIysIle 160  
 Db 586 GATCTCAAGAGCCATTTGACAAAGCCCTGGAAGATTTATGAGCGAAGTTTACAAAAAT 645  
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
 Db 646 GAGAGGAGAAAGAGACACGCAAAACAGACCGGATGATCCGACAGAGATACAGGC 705  
 Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
 Db 706 GCTGAGATTGAGAGAAATGGAAGAAAGGAAACGGCGCTTTTTCAGCTCCAGATGTGTA 765  
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220  
 Db 766 TATCTCAATTAAGTAAATGAATCAAGACCAAAAGGGGTGTGGACCTCTGCAGAACCTT 825  
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
 Db 826 ATAAAGTATTACCATGCAAGTGCAATTTCTTTTCAAGATGGCTTGAAACCGCTGTATAA 885  
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 Db 886 CTGAAACAGTACATAGAAAGCTGCTGCTGACTTATATATATATAATAAAGCAACCAAGAT 945  
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
 Db 946 GAAGAAAAAACAACAGCTAACTGCACTCCGAGATCTTAATAAATCTCTCTGCAACTTGA 1005  
 Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 Db 1006 CCGAAGAGAGACTCTCAGAGCGGCAAGGT---GGGTACAGCATGCATCAGCTCCAGGCG 1062  
 Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320

Db 1063 AACAAAGGAGTACGGCAGTGAAGAGGGCTTCTGCTGAAGAAAGAGTGTGGATCCGG 1122  
 Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
 Db 1123 AAAGTGTGGCAGACAGCGAGGTGTCGGTCAAGAACGGGATCTCCACGCCA 1182  
 Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
 Db 1183 ACTTCCAACAGCGACCTGCTAAGTTAAACCTCTACCTGCGCAAGTGAACCAAGTGT 1242  
 Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
 Db 1243 GAGCACAAGAAGTCTTTTGACCTGATATCATATACATATCGAATCATATCTTTCAAGCAGAA 1302  
 Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
 Db 1303 GATGAGCAGGATATATAGCGTGGATATCAGTACTGAGAAATAGCAAGAGAGCCCTA 1362  
 Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
 Db 1363 ACCATGGCTTCGCTGCTGAGCAGCAGCAGGGGAGAGATAGCCTG---GAGGACCTTACC 1419  
 Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
 Db 1420 AAAGCCATCATCGAGGATGTACAGCGCTCCCTCGGAATGACATCTGCTGTACTGGCGC 1479  
 Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
 Db 1480 TCATCAGAACCACCGTGGCTTTCAACCAACTTGGGTATTTTGACCTGTATAGAATGTTC 1539  
 Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
 Db 1540 GGAATCCATAGGAAATGGGGTTTCATATTCTCGCATTCAGTCTTTTGGAACTAGACAAA 1599  
 Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
 Db 1600 TTAGAACTTCGAACTCTTCTGCGCAAGAAATAGGAAACAATAGTATTTTAATGATATT 1659  
 Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
 Db 1660 ATGAAGCAATTTTACCAGC---CCTTCAACAAAGCCACCCCTTCAAGTGATATGACC 1716  
 Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
 Db 1717 GTGGCTTAAGAGTACATCACTGCAAGATGTCTCGACCCGCTTCTCCAGGAGACCTGT 1776  
 Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
 Db 1777 GCCTTCTTCTCAGCAAGCTGAACGAATTTGCTTGAGGCCATCAAGTCCAGGACTTACTT 1836  
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
 Db 1837 GCATTAATCAAGTCTATGACAGGGGTGGAGTTAATGGAGCCACTG---CTGGAACCT 1893  
 Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
 Db 1894 GGGCAGAGCTCGGGGAGACAGCCCTTCACTTCCCTCAGGACCGCAGACCATCT 1953  
 Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
 Db 1954 CTCATCTGGTTGATTTCTTTGTACAAACTGTGGAACTGTGGAACTGGGATAAGCAACCTGTGTG 2013  
 Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
 Db 2014 GGAATACAGTCTGCACTATTGACAGTGTACCGCAAGCCGGAGTGTCTGAAAGCTGTCTG 2073  
 Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
 Db 2074 CTTAGGAGCAAGCCACCGTGACATCGTTAACAGAGATGGAGAACTGCCCTTGGATATA 2133  
 Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
 Db 2134 GCAAGAGACTCAAAAGCTACCCAGTGTGAAGATCTGCTTTCCAGGCTAAATCTGGAAG 2193





LAKVGNNSFNDIMEANLPSPPKPTPSSDMTVRKEYITAKYVVDHRRSRKTCSSSSAK  
 FLNELLEATKSDLLALIOVYAYGVELMEPLPGQELGETALHLAVRADTQDLSHLVD  
 LFNQNCNGLDKQTALGNTHALHYCSMYKPECLKULLRSKPTVDVUNQAGETALDIAXR  
 LKATQCBLLLSQAKSKCNPHVVEYNLROBEMDESDDLLDKPSFKIKERSPRPQ  
 SFCHSSSISPODKLSLPGFSTPRDKORLSYGAFNQIFVSTSDSPSPTPIAEAPPLPP  
 RNATKGGPPSTLPLSTQTSFGSSTLSKRSPPPPGPKRTLSDDPSPLPHGPPNKG  
 AVPWGNDVGPSSSKTWNKFGLSQOQSSTGSAKALGRVLPKLPQKVALKRTSTSHH  
 LSLDKANVPPIPIFOKSSOLLTELPOKPPGDLPPKPTTELAAPKPTGLDLPKPGELPPKP  
 QLGLDLPKPEQLADLPPKQVXDLPPKPOLGELLAKPQTGDASPKPAQPPLELTPKSHPA  
 DLSNPVKPQASEDINDLPTLPETPLPRKINTGSKVRRVKTIYDCQANDDELTF  
 MEGEVIVVTGSEDOEWMTGHIEGQPERKGVFPVSVFHTLSD\*

## ORIGIN

Alignment Scores:  
 Pred. No.: 9,77e-104 Length: 5330  
 Score: 3013.50 Matches: 627  
 Percent Similarity: 66.67% Conservative: 129  
 Best Local Similarity: 55.29% Mismatches: 225  
 Query Match: 57.15% Indels: 153  
 DB: 4 Gaps: 23

US-09-914-042-1 (1-1006) x AF112886 (1-5330)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAla 20  
 Db ATGCCGGACCAAGATCTCCGTCTCCGAGTTCATCGCGGAGACACCGAGGACTACAACTCG 328  
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
 Db CCCACCAGCTCCAGCTTCACTACGGGGCTGCACAACTGCAGGAACACCGTCAACGCTGCTG 388  
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
 Db GAGGAGGGCTCTAGACCAAGATAGAACACGCTTACAGAAAGTTAAGAGTCTGTAAAGCA 448  
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80  
 Db ATATACAAATTCGGTCAAGACCATGTACAAATGAAGAAACTATGCGCAAGTCTTGTAT 508  
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
 Db AAGTTTGGGAGTAATTTTTTAAGTCGAGACACCCAGATCTTTGGCACCCTTTTGTCAAG 568  
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 Db TTTTCTACGCTTACAAAGAACTGTCCACACTGCTGTAATAAATCTCTCCAGGGCTGAGC 628  
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140  
 Db CACAATGTGATCTTCACCTTGGATTCCTTGTGTAAGGAGAGACTGAAGGAGTCAAGGC 688  
 Qy 141 AspLeuLysLysPhePheAspLysAlaThrLysAspTyrGluThrLysIleThrLysIle 160  
 Db GATCTCAAGAAACCAATTTGACAAAGCTTGGAAAGATTATGAGCAAGATTTTACCAAAAT 748  
 Qy 161 GluLysGluLysLysGluHiAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
 Db GAGAAGGAGAGAGGAGGAGCACCAGCAGCAGGAGATGATCCGACCGAGATCAACCGC 808  
 Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
 Db GCCGAGATCGCGAGGAAATGAAAGGAGCGCGCTCTTCCAGCTCCAGATGTGCGAG 868  
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
 Db TATCTCAATTAAGATTAATGAATCAAGACCAAAAGGGGTGTGGATCTGTGCGAGAACCTG 928  
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
 Db ATAAAGTATTATCACGCAGTGCATTTCTTTCAAGATGGTTTGAAGACAGCTGATAAA 988  
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 Db

Db 989 TTGAAACAGTACATTGAAAGCTGGCTGCTGATTGTGTATATATCAAAACAGACCCAGGAC 1048  
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGluValGlu 280  
 Db 1049 GAAGAAAGAAACAGCTGACCGCACTCCGAGACCTAATAAAGTCTCGCTCCAACCTCAT 1108  
 Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
 Db 1109 CAGAAGGAGTCTAGAGAGATTCAGAGCGCGGAGGGA---GGCTACAGCATGCACCAG 1165  
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
 Db 1166 CTGACGAGCAACAAGGAATACGCGCAGAGAAAGGGCTACCTGCTGAAGAGAGTGCAC 1225  
 Qy 318 GlyIleArgLysValTropGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
 Db 1226 GGGATCCGGAAGTGTGCGAGAGAAAGAGTCTCCGTCAAGAACGGGATCCTGACCAATC 1285  
 Qy 338 SerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
 Db 1286 TCCACAGCCAGTCCACAGACAGCCAGCCAGCTGAACCTTCTCACTTGCAGGAGTGAAG 1345  
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377  
 Db 1346 CCGAATCCGAGGACAAAGAGTCTTTTACCTGATATACATAACAGAGACGTATACACTTT 1405  
 Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTropMetSerValLeuGlnAsnSerLysGlu 397  
 Db 1406 CAGGCCCAAGATGACAGGATTTATGTACGCTGGATCTCGGTGCTGACAAAACAGCAAGAG 1465  
 Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417  
 Db 1466 GAGGCCCTTCACCATGGCTTCCGGGGGAGACAGAGTCTGGGGAGAGAGCGCTG---GAG 1522  
 Qy 418 GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
 Db 1523 GAGCTGACGAAGGCATCATCGAGAGCTGCAGCGGTCCCGGGCAACAGAGCTCTGTGTC 1582  
 Qy 438 AspCysGlyAlaProAspProThrTropSerThrAsnLeuGlyIleLeuThrCysIle 457  
 Db 1583 GACTGCGGCTCGGAGAACCCACCTGCTGTCACCAACTTGGGCATCTTGACCTGTATA 1642  
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
 Db 1643 GAATGTTCCGCATCCATAGAGAAATGGGGTTCATATTTCTCGCATCCAGCTCTTGGAA 1702  
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
 Db 1703 CTAGACAAATTAGGAACCTTCTGAACCTCTGCTGCGCAAGAAATGTAGGAAAACAATAGTTT 1762  
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
 Db 1763 AATGATATTATGGAAGCAAAATTTTACCAGT---CCCTCACCAAAACCCACCCCTTCAAGT 1819  
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArg 537  
 Db 1820 GATATGCTGTACGGAAGGAATATATCACTCAAGATATGTAGATCATAGGTTTTCACGG 1879  
 Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
 Db 1880 AAGACCTGTTTCATCGTCATCAGCTAAACTGAAGTAACTGCTTGGAGCCATCAAAATCCAGG 1939  
 Qy 558 AspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
 Db 1940 GATTACTTGCATTAATTCAGTCTATGACAGGGGGTGGAGCTAATGGAACCGCTG--- 1996  
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
 Db 1997 CTGGAACCCGACAGGAGCTTGGGAGACACCCCTTCATCTTGCAGTCCGAAACCCGACAG 2056  
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
 Db 2057 CAGACATCTCTCCATTTGGTGGACTTCTCTGTACAAAACCTGTGGGAACCTAGATAAGCAG 2116



Db 389 GAGGAGGCTCTAGACCAAGATAGAACAGCGCTTACAGAAAGTTAAGAAAGCTGTGTAAGACA 448  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyThrGlnAlaLeuGlu 80  
Db 449 ATATACAAATCCGGTCAAGACCATTGTACAAATGAAGAAACTATGCCAAAGTTCTTGAT 508  
Qy 81 LysPheGlyGlyAsnCyValCyArgAspProAspLeuGlySerAlaPheLeuLys 100  
Db 509 AAGTTTGGAGTAATTTTAAAGTCGACGACCAACCCAGATCTTGGCCCGCTTTTGTCAAG 568  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db 569 TTTTCTACGCTTACAAAGAACTGTCCACACTGTCTGAAATAATCTGCTCCAGCGCGCTGAGC 628  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 629 CACAATGTGATCTTCCACTTGGATCTCTTGTGAAGGAGAGCCTGGAAGGAGTCAAAAGGC 688  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 689 GATCTCAAGAAACCAATTTGACAAAGCTTGGAAAGATTATGACGAGATTTACCAAAATT 748  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db 749 GAGAAGGAGAGAGGGAGCACGCCAAGCAGCAGCGGATGATCCGACGAGATCACCGGC 808  
Qy 181 AlaGluIleAlaGluLysGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
Db 809 GCCGAGATCGCGGAGGAATGGAAAGGAGGCGCGCTCTTCCAGCTCCAGATGTGCGAG 868  
Qy 201 TyrLeuLeuLysValAsnGluLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 869 TATCTCATTAAGTTAATGAATCAAGACCAAAAGGGTGTGGATCTCTGTCGAGAACCTG 928  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 929 ATAAAGTATTATCACGCACAGTGCATTTCTTCAAGATGGTTTGAAACAGCTGATAA 988  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 989 TTGAACACAGTACATTGAAAGCTGGCTGCTGATTTGTATATATATCAACACAGACCAGAC 1048  
Qy 261 GluGluArgArgGlnLeuGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1049 GAAGAAAGAAACACAGCTACCGCACTCCGAGACCTTAATAAAGTCTCTCTCCACTCGAT 1108  
Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
Db 1109 CAGAAGGAGTCTAGGAGAGATTCCACAGCGCGGACGGGA---GGCTACAGCATGCACCAG 1165  
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
Db 1166 CTGAGGCGCAACAGGAATACGGCAGCGAGAGAAGGGCTACTCTCTGAAGAAGAGTGAC 1225  
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCySerValLysAsnGlyPheLeuThrIle 337  
Db 1226 GGGATCCCGAAGTGTGCGAGAGAAGAGTGCTCCGTCAGAACCGGATCTCTGACCATC 1285  
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuLeuThrCysGlnValLys 357  
Db 1286 TCCACGCCACGCTCCACACAGACAGCCAGCTGAACCTTCTCACTTGCCAGGTGAAG 1345  
Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377  
Db 1346 CCGAATCGCGAGGACAAGAAGTCTTTTGACCTGATATCACATAACAGACGATATCACCTT 1405  
Qy 378 GlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
Db 1406 CAGGCCGGAAGATGACAGAGTATTAGCTGGATCTCTGGTGCTGACAAACAGCAAGAG 1465  
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGln 417  
Db 1466 GAGGCGCTCACCATGGCTTCGGGGGGGAACAGAGTGTGGGGAGAGCAGCTG---GAG 1522

Qy 418 GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
Db 1523 GAGCTGACGAAGGCCATCATCGAGAGGTGCGAGGGCTCCGGGCAACGAGCTGTCTGTC 1582  
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457  
Db 1583 GACTTCGGCTCGGCAGAACCCACCTGGCTGTCCACCAACTTGGGCACTTTCACCTGTATA 1642  
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
Db 1643 GAATGTTCCGCGATCCATAGAGAAATGGGGTTTCATATTTCTCGCATCCAGCTTTGGAA 1702  
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
Db 1703 CTAGACAAATTAGGAACCTTCTGAACCTCTGTGCGCAAGATGTAGGAAACAATAGTTTT 1762  
Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
Db 1763 AATGATATTAGGAAGCAAAATTTACCCAGT---CCCTCACCACCAACCCACCTTCAAGT 1819  
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537  
Db 1820 GATATGACTGTACGGAAGGAATATATCACTGCAAAAGTATGTAGATCATAGGTTTTCACGG 1879  
Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
Db 1880 AAGACCTGTTTCATCGTCATCAGCTAAACTGAACGAATGCTTGAGGCCATCAAAATCCAGG 1939  
Qy 558 AspilePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
Db 1940 GATTTACTTGCACTTAATTCAGTCTATGCAGAGGGGTGGAGCTAATGGAACCCGCTG--- 1996  
Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
Db 1997 CTGGAACCCGACAGGAGCTTGGGGAGACAGCCCTTCATCTTGCAGTCCGAACCGCAGAC 2056  
Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
Db 2057 CAGACATCTCTCCCAATTTGGTGGACTTCTCTGTACAAACTGTGGGAACCTAGATAAGCAG 2116  
Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeu 637  
Db 2117 ACGGCCCTGGGGAACACGGCCCTGCACCTACTGTAGTATGTACAGTAAACACAGAGTGTG 2176  
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657  
Db 2177 AAGCTGTCTCTCAGGACAGCCCACTGTGGACGTCTGTAATCAGGCTGGAGACCGCC 2236  
Qy 658 LeuAspileAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677  
Db 2237 CTGACATAGCAAGAGACTGAAAGCCACTCAGTGTGAAGACCTGTCTTCCAGCTAAA 2296  
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697  
Db 2297 TCTGGAAGTTCAATCCACAGCTCCACGTGGAATATGAGTGAATCTTCGACAGGAGAG 2356  
Qy 698 LeuAspGluSerAspAspMetAspMetAspGluLysLeuGlnProSerGluAsnArgGlu 717  
Db 2357 ATGGATGAGCGATGACGACCTGGATGACAAACCGAGCCCGCCCATCAAGAGGAGCGCTCC 2416  
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737  
Db 2417 CCCCAGCCGAGAGCTTCTGCGCACTCTCCAGCATCTCCCCCAGGAC-----AAG 2467  
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGlnArgAlaPheMetProSerIle 757  
Db 2468 CTCTCAGCTCCGGGCTTCAGCACCGCCCAAGGCAAGCAACGACTC----- 2512  
Qy 758 LeuGlnAsnGluThrTyrGlyAla-----LeuLeuSerGlySerPro 771  
Db 2513 -----TCCTACGGCGCTTCACCAACACAGATCTTCGTCTCCACAGCAGACAG 2560

Qy	772	ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsn	791
Db	2561	TCACCCACGTCACCGATCGCA-----GAGCGCGCCGCTGCTCCCAAGAAC	2608
Qy	792	ValGlyLys-----ValGlnThrAlaSer	799
Db	2609	GCCACGAAGGTCCACCTGGCCCACTTCAACACTCCCTCTAAGCACCCAGACTCTVAGT	2668
Qy	800	SerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArg	819
Db	2669	GGCAGCTCACCCCTGTCCAAGAAGCGGTCTCTCCCTCCCAACCCAGGCGAGTTCCT	2725
Qy	820	SerSerSerAspProProAlaValHisProProLeuProProLeuArg-----ValThr	837
Db	2726	ACCTGTCTGACCTCCACGCCCACTACCTCAGGGCCCCCAACAAGGCGAGTTCCT	2785
Qy	838	SerThrAsnProLeuThrProThrProProProValAlaLysThrProSerValMet	857
Db	2786	TGGGGTAAACAGTGGGTCTCTCATCTGCC-----AGTAAGACCAACGAACAAGTTC	2836
Qy	858	GluAlaLeuSerGln-----	862
Db	2837	GAGGGCTGTCCCGACGAGTCGAGCACCGGTCTGCAAGACTGCATTTGTCCCAAGATT	2896
Qy	863	-----ProSerLysProAla-----	867
Db	2897	CTTCTAAACTACCTCAGAAAGTGGCATAAGGAACACAGACCAACCATCATCTCTCC	2956
Qy	868	-----ProProGlyIle-----SerGlnIle-----	874
Db	2957	CTCGACAAAGCCAAACGTCCCACTGAGATCTCCAGAAGTGTGCCAGTTTGACAGAGTTA	3016
Qy	875	-----ArgProProPro-----LeuProProGlnPro-----	883
Db	3017	CCGAGAAAGCGCCACCCGGGACCTGCCCCCGAAGCCACAGAACTGGCTCCCAACCC	3076
Qy	884	-----ProSerArgLeuProGlnLys-----	890
Db	3077	CCCATTTGGAGACTTACCACCTAAGCCAGCGAGCTGCCCCCGAAGCACAGCTGGCGCAG	3136
Qy	890	-----	890
Db	3137	CTGCCCCCAAGCCCCAGCTCGCAGACTTGCCCCCAAGCCCCAGGTGAAGACCTGCCT	3196
Qy	891	-----LysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg	909
Db	3197	CCCAAGCCCAACTGGGGGAGCTGTGGCAAAACCCACAGCGGAGAGCGCTCGCCCAAG	3256
Qy	910	Gly-----ProValAspLeuSer-----	915
Db	3257	GCCAGCACCCCTGGAGCTCACCCCAAGTCACACCCGGCGGACCTGTCCCCGAAGCTC	3316
Qy	916	-----AlaThrGluAlaLeuGlyProLeuSerSerAsnAlaMetValLeuGlnProPro	932
Db	3317	CCCAAGCAGCGCTGTAGGACACCAACGACCTCACGCCCAACCCCTGCCA---GAGACACCC	3373
Qy	933	AlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeu	952
Db	3374	GTGCCTCTGCCAGGAAGATCAACACGGGGAAGACAAAGTGTAGGGCGAGTGAAGACCATC	3433
Qy	953	TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIle	972
Db	3434	TACGACTGCCAGGGGACACAGATGACGACCTGACTTTTCATGAGGGCGAGGTGATCGTG	3493
Qy	973	ValAspGlyGluGlnAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArg	992
Db	3494	GTACCCGGGAGGAGGACAGGAGTGTGTATTGGGCACATCGAGGGGACGCCGAGAGG	3553
Qy	993	LysGlyAlaPheProValSerPheValHisPheIleAlaAsp	1006
Db	3554	AAGGGCGTCTTCCAGTGTCTTTGTCCATCATCTGTCTCGAC	3595

RESULT 15

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20

AF075462 4150 bp mRNA linear ROD 28-DEC-1998  
LOCUS Mus musculus ADP-ribosylation factor-directed GTPase activating  
DEFINITION protein isoform b (Shagl) mRNA, complete cds.  
ACCESSION AF075462  
VERSION  
KEYWORDS AF075462.1 GI:4063615  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4150)  
AUTHORS Brown, M.T., Andrade, J., Radhakrishna, H., Donaldson, J.G.,  
Cooper, J.A. and Randazzo, P.A.  
TITLE ASAP1, a phospholipid-dependent arf GTPase-activating protein that  
associates with and is phosphorylated by Src  
Mol. Cell. Biol. 18 (12), 7038-7051 (1998)  
JOURNAL 990338209  
MEDLINE 9819391  
PUBMED  
REFERENCE 2 (bases 1 to 4150)  
AUTHORS Brown, M.T., Andrade, J., Radhakrishna, H., Donaldson, J.G.,  
Cooper, J.A. and Randazzo, P.A.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1998) Laboratory of Cellular Oncology, National  
Cancer Institute, 37 Convent Dr. MSC 4255, Bethesda, MD 20892, USA  
FEATURES  
source 1..4150  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
gene 1..4150  
/genes="Shagl"  
CDS 50..3322  
/genes="Shagl"  
/notes="SHAG1b; ASAP1; alternatively spliced"  
/codon\_start=1  
/product="ADP-ribosylation factor-directed GTPase  
activating protein isoform b"  
/protein\_id="AAC98350.1"  
/db\_xref="GI:4063616"  
/translation="MRSSASRLSFSRSDLSLNRMPDQISVSEFTAEETDYSPTTS  
STFTRHNCRTVTLLEAEDQRTALQKVKSVKAIYNSGDHQQNENYIAQVLDKF  
GSNFSRNPDLTGFVFKFTIKELSTLLKQLGSLHNVIPTDLSLKGDLKGVKG  
DLKQFPDKAWQDYETKFTIKEREKREHAKQHGMIRTEITGAEIAEEMEKERLLFQMG  
CEVLIKNVEIKTKGVDLLNLIKYVHAQCNFFODGLKTADKLKQYIEKLAADLVNLIK  
QTQDEKKQLTALDLIKSLQLDPKEVGLVVASRANSRDSRSQSGYSMHQLOK  
NKEYSGEKGFLFKGDLGIRKQWRKCAVNGILTISHATSNRQPAKUNLLTCQVKP  
NAEDKRSFLLSHNTHYFQAEDQDYTAWISVNSKEEALTMAPRGESGTGENSL  
DUTKAIIEVDQLPGNDICDCGSEPTWLSNLGILTCIEGSIHREMGVHISRIQS  
LELDLQTELLKANNVNSNDMEANLPSPKPTPSDMTVRKEYITAKYVDHVR  
FSRQTCASSAKLLELLEIKSRDILLALIQVAEGVELMEPIELPGQELGETALHLAV  
RTAQTSLHLVDFVQNCNLQKTSVGNTVLHYCSMYGKPECLKLLLRKSTPVDIV  
QNGETALDIAKKLTKQCEBLLLSQAKGFPNPHVVEYEWNRQDESDSDDDLDLQK  
STIKERSRPPQSFCSSISISPDKLALPGFPRDKQLRSYGFATNQIFASTDPLP  
TGPTGAPPLPRHAKGNDVPSKSTANKFEGLSQAQSTSAKALTALGPRVLPKLP  
QKVALRKTTSHSLDRNTNIPETFOKSSOLTTELPOKPLGELPPKPVELAPKQVQ  
ELPPEGELPPKQGLDLPKQSDLPKPKQMDLPKPKQGLLAKSQAGDVSNKY  
QPQSVTYQKSTGDSLPVQSRDAIQKSDSESDNDLTPLPETPVPLEPKINIGKNV  
RRVVTIYDQADNDDELTFIEGVEIIVTGEEDQEWIIGHIEGQRPVFPVFSVHLL  
SD"

ORIGIN

Alignment Scores:	9,34e-104	Length:	4150
Pred. No.:	3011.00	Matches:	617
Score:	68.10%	Conservative:	128
Percent Similarity:	56.40%	Mismatches:	237
Best Local Similarity:	57.10%	Indels:	112
Query Match:	10	Gaps:	18
DB:			

US-09-914-042-1 (1-1006) x AF075462 (1-4150)

Alignment Scores:		
Pred. No.:	9.34e-104	Length:
Score:	3011.00	Matches:
Percent Similarity:	68.10%	Conservative:
Best Local Similarity:	56.40%	Mismatches:
Query Match:	57.10%	Indels:
DB:	10	Gaps:
US-09-914-042-1 (1-1006) x AF075462 (1-4150)		

```

Db      |||||||ATCCCGGACAGATCTCGGTGCGAGTTTCATCGCGGAGACCACCGAGGACTACAACTCG 169      :
Qy      |||||||ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40      :
Db      |||||||CCCACCAAGTCCAGCTTCACACCGCGGTGCACAACTGCAGGAAACACCGTCCACGCTGCTG 229      :
Qy      41  GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60      :
Db      |||||||GAGGAGGCTCTAGACCAAGATAGAACAGCCCTACAGAAAGTGAAGAAATCTGTAAGACA 289      :
Qy      61  IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80      :
Db      |||||||ATATAATATCCGGTCAAGATCATGTACAAATGAAGAAACTATGCACAACTGTCGAT 349      :
Qy      81  LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100      :
Db      |||||||AAGTTTGAAGTAATTTTAAAGCAGACAAACCTGACCTTGGCACCCTTTTGTCAAG 409      :
Qy      101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120      :
Db      |||||||TTTTCTACACTCAAAAGGAATGTCACGCTGCTGAAAAATCTGCTCCAGGGTCTGAGC 469      :
Qy      121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140      :
Db      |||||||CACAAATGTGATCTTCACCTTGATTCCTTGTAAAGGAGAGACCTGAAGGGGGTCAAGGA 529      :
Qy      141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160      :
Db      |||||||GATCTCAAGAGCCATTTGACAAAGCTTGAAGAGATTATGAGACGAAATTTACAAAAATT 589      :
Qy      161 GluLysGluLysLysGluHisAlaLysLysGlyMetIleArgThrGluLysSerGly 180      :
Db      |||||||GAGAGGAGAAAGAGAGACGCAAAACAGACCGGGATGATCCGAACAGAGATTAACAGGC 649      :
Qy      181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200      :
Db      |||||||GCTGAGATTGAGAGAAATGGAAAAAGAACGGCGGCTTTTTCAGCTCCAGATGTGTGAA 709      :
Qy      201 TyrLeuLeuLysValAsnGluLysIleLysLysLysGlyValAspLeuLeuGlnAsnLeu 220      :
Db      |||||||TATCTCATTAAGTAAATGAAATCAAGACCAAAAGGGGTGTGGACCTGCTGCAGAACCTT 769      :
Qy      221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240      :
Db      |||||||ATAAGTATTACCATGCACAGTGAATTTCTTCAGGATGGCTTGAAACCGGCTGATAAA 829      :
Qy      241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260      :
Db      |||||||CTGAACAGATACATAGAAAGCTGCTGCTGACTTATATATATAAAGCAACCAACAGAT 889      :
Qy      261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280      :
Db      |||||||GAAGAAAAAAACAGCTAACTGCACTCCGAGATCTTAATAAAATCCTCTCTGCAACTGAT 949      :
Qy      281 GlnLysGlu-----AspSer 285      :
Db      |||||||CCGAAAGAGTAGTGTTTATATGTTCCAGCAGCGGCTAACAGTTCTAGAAGAGACTCT 1009      :
Qy      286 GlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlnLysLysGluHisGly 305      :
Db      |||||||CAGACCCGGCAAGGT---GGGTACAGCATGATCATGCTCCAGGGCAACAGAGATACGGC 1066      :
Qy      306 ThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLys 325      :
Db      |||||||AGTGAAGAAAGGGGCTTCCTGCTTAAAGAGTGTGATGGATCCGGAAGTGTGGCAGAGA 1126      :
Qy      326 ArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgPro 345      :
Db      |||||||CGAAGTGTGGGTCAAGAACGGGATCTCTGACCATCTCCACGCAACTTCCAACAGGCAG 1186      :
Qy      346 ProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGluGluLysLysCys 365      :

```

```

Db      1187 CCTGCTAAGTTAAACCTCTCTCACCTGCTCAAGTGAACCCGAATGCTGAGGACAAGAGTCT 1246
Qy      366 PheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCys 385
Db      1247 TTTGACCTGATATACATAATCGAACAATATCATTCTCAAGCAGAAGATGACGAGATTAT 1306
Qy      386 GlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLys 405
Db      1307 ATAGCGTGGATATCAGTACTGACGATAGCAGAGAGGCCCTTAACCATGGCCCTTCGT 1366
Qy      406 GlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleSer 425
Db      1367 GGTGAGCAGAGCACAGCGGAGAAATAGCCTG---GAGGACCTTACCAAGCCATCATCGAG 1423
Qy      426 GluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThr 445
Db      1424 GATGTACAGCGGCTCCCTGGGAATGACATCTGCTGTGACTGCGGCTCATCAGAACCCACG 1483
Qy      446 TrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGlu 465
Db      1484 TGGCTTTCAACCAACTTGGGTATTTTGACCTGTATAGATGTTCCGGAATCCATAGGAA 1543
Qy      466 LeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGlu 485
Db      1544 ATGGGGTTTCATATTTCTCGCATTCAGTCTTTTGAACCTAGACAAATTAGGACTTCGAA 1603
Qy      486 LeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysLeu 505
Db      1604 CTCTTGTGTCGCAAGATGTAGGAACAATAGTTTAAATGATATATTAGGAAGCAAAATTTA 1663
Qy      506 ProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyr 525
Db      1664 CCCAGC---CCTTCACCAAAAGCCACCCCTTCAGTGATATGACCGTCCGTAAAGATAC 1720
Qy      526 IleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAla 545
Db      1721 ATCACTCAAAAGTATGTCGACCACCGCTTCTCCAGGAAGACCTGTGCGCTCTCTCCACGA 1780
Qy      546 LysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAla 565
Db      1781 AAGCTGAACCAATTTGCTTGAGGCCATCAAGTCCAGGACCTTACTTGGATTAAATCCAAGTC 1840
Qy      566 TyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAsp 585
Db      1841 TATCCAGAGGGGTGGATTAATGAGCCACTG---CTGGNAACCTGGCGAGACTCGGG 1897
Qy      586 GluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAsp 605
Db      1898 GAGACAGCCCTTCATCTTGGCGTCAGACCGCAGACAGACATCTCTCCATCTCGTTGAT 1957
Qy      606 PheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeu 625
Db      1958 TTCCTTGTACAAAACCTGTGGGAACCTGGTAAGCAAAACGCTGTGGGGAATACAGTTCTG 2017
Qy      626 HisTyrCysCysLeuThrAspAsnAlaGlyCysLeuLysLeuLeuArgGlyLysAla 645
Db      2018 CACTATTGACAGTGTACGGCAACCGAGTGTCTGAAGCTGTCTGTAGGAGCAAGCCCC 2077
Qy      646 SerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLys 665
Db      2078 ACCGTGGACATCGTTAACCAAGATGGAGAACTGCCTTGGATATAGCAAGAGACTCAAA 2137
Qy      666 HisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisVal 685
Db      2138 GCTACCCAGTGTGAAGATCTCTTCCAGGCTAAATCTGGAAAGTTCAACCCATCATGTC 2197
Qy      686 HisValGluTyrGluTyrArgLeuLeuHisLeuAspLeuGluSerAspAspAspMet 705
Db      2198 CACGTAGATGATGATGGAATCTTCGACAGACAGATGGATGAGAGCATGACGATCTG 2257
Qy      706 AspGluLysLeuGlnProSerGluAsnArgGluAspArgProIleSerPheTyrGln 725
Db      2258 GATGACAAAGCCAGCCGATCAAAAAGGAACCGCTCGCCACAGACCACAGAGCTTCTGCCAC 2317

```

Search completed: August 4, 2005, 11:54:31  
Job time : 10750 secs

```
Qy 726 LeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeu 745
Db |||||
2318 TCCTCCAGCATCTCGCCACAGGAC-----AAGCTGGCACTGCCAGGTTTCAGCACT 2368
Qy 746 AlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAla 765
Db :::::|||||
2369 CCACGGGACAAAGCAGCGGCTC-----TCCTACGGAGCC 2401
Qy 766 Leu-----LeuSerGlySerProProAlaGlnProAlaAlaProSerThrThr 782
Db |||||
2402 TTCACCAACCAAGATCTTCGCCTCTACGAGCAGAGATTGGCCACATCACCC---ACCAGT 2458
Qy 783 SerAlaProProLeuProProArgAsnValGlyLys-----ValGlnThrAlaSer 799
Db |||||
2459 GAGGCTCCCTTTGCGCACCTCGGAACCGCGGAAAGGTAATGATGTGGGCCCATCATCT 2518
Qy 800 SerAlaAsnThrLeuThrLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819
Db |||||
2519 TCAAGTAAGCGCAACAAAGTTTGAGGGGTGTCTCAGCAAGCAAGCACCAGTCTTGCT 2578
Qy 820 SerSerSerAspProProAlaValHisProLeuPro-----ProLeuArgVal 836
Db :::::
2579 AAGACTGCCCTTGGCCCGAGAGTGCTTCTTAACCTCAGAAAGTGGCACTAAGGAAG 2638
Qy 837 ThrSerThrAsn----- 840
Db |||||
2639 ACGGAGACCAAGCCATCATCTCTCCCTCGCAGACCAACACATCCACCTTGAGACTTTTCAG 2698
Qy 841 ----- 844
Db |||||
2699 AAATCATCACAGTTGACAGATTACCCCAAGGCCACCACTTGGAGAGCTGCCCGGAAG 2758
Qy 849 ProVal-----AlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys--- 865
Db |||||
2759 CCTGTGGNACTGGCCCCCAAGCCCAAGTTGGAGAGCTGCCACCTAAGCCTGGAGAGCTA 2818
Qy 866 -----ProAlaProProGlyIleSerGlnIleArgPro 876
Db |||||
2819 CCCCTAAGCCCCAATTAGGTGACCTGCCCCCAAGCCACAGCTCTCAGACTTACCTGCC 2878
Qy 877 ProPro-----LeuProPro----- 881
Db |||||
2879 AAGCCACACAGATGAAGACCTGCCCTTAAGCCGAGCTGGGGGATCTGCTGGCAAGTCC 2938
Qy 882 -----GlnProProSerArgLeuProGlnLysLysPro 892
Db |||||
2939 CAGGCTGCGGATGCTCAGCCAAGGTGCGCCACCTCCCAATGTACAGTCCAGAGATGCCATC 2992
Qy 893 AlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProVal 912
Db |||||
2993 -----TCACACACCGGGATCTGTCTCCAAATGTACAGTCCAGAGATGCCATC 3040
Qy 913 AspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProPro 932
Db |||||
3041 CAGAAGCAAGATCTGAAGACTCCAAACGACCTCACACCCAGCTGCCA---GAGACACCC 3097
Qy 933 AlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeu 952
Db |||||
3098 GTACCACTGCCGAGAAAATCAATACGGGGGAAATAAAGTGAAGCGGGTGAAGACCAATT 3157
Qy 953 TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIle 972
Db |||||
3158 TATGACTGCCAGGCAGATAATGACGACCACTCATATTTATTGAGGGGGAAGTATCATTT 3217
Qy 973 ValAspGlyGluGluAspGlnGluThrTrpIleGlyHisIleAspGlyAspProGlyArg 992
Db |||||
3218 GTCACCGGGGAAGAGGACAGGAGTGTGTGATCGGGCATATCGAAGGACAGCCTCAAGG 3277
Qy 993 LysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db |||||
3278 AAGGTGTCTTTCCAGTGTCTTTTGTCCACATCTCTGTCTGAC 3319
```

***THIS PAGE BLANK (USPTO)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:26:55 ; Search time 103 Seconds  
(without alignments)  
3813.012 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHEDYKA.....DGDPRKGAFPVSVFHIAD 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1752860 seqs, 390397842 residues

Total number of hits satisfying chosen parameters: 1752860

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3006	57.0	1132	15	US-10-210-281-74
2	2071.5	39.3	903	15	Sequence 74, Appl
3	2071.5	39.3	903	17	Sequence 2951, Ap
4	1801.5	34.2	350	11	US-10-490-605-2
5	1759.5	33.4	349	11	Sequence 2, Appli
6	1075.5	20.4	208	15	Sequence 635, App
7	611.5	11.6	358	16	Sequence 956, App
8	569.5	10.6	764	15	Sequence 2284, Ap
9	558	10.4	804	15	Sequence 2815, Ap
10	551	10.4	804	15	US-10-080-334-238
11	533	10.1	759	15	Sequence 80, Appl
					Sequence 235, App
					Sequence 236, App

12	529	10.0	834	14	US-10-176-306-11	Sequence 11, Appl
13	529	10.0	834	15	US-10-080-334-237	Sequence 237, App
14	529	10.0	834	15	US-10-467-434-4	Sequence 4, Appli
15	509.5	9.7	775	16	US-10-739-930-6184	Sequence 6184, Ap
16	509	9.7	740	15	US-10-080-334-239	Sequence 239, App
17	487.5	9.2	776	16	US-10-729-930-6289	Sequence 6289, Ap
18	480	9.1	156	15	US-10-276-774-1638	Sequence 1638, Ap
19	438.5	8.3	773	16	US-10-437-963-136533	Sequence 136533,
20	419	7.9	605	9	US-09-828-303-19	Sequence 19, Appl
21	419	7.9	605	16	US-10-716-089-19	Sequence 19, Appl
22	408.5	7.7	503	11	US-09-833-245-1843	Sequence 1843, Ap
23	406	7.7	136	15	US-10-276-774-1754	Sequence 1754, Ap
24	398	7.5	324	17	US-10-921-707-3	Sequence 3, Appli
25	396.5	7.5	732	16	US-10-437-963-181247	Sequence 181247,
26	391.5	7.4	649	16	US-10-437-963-181939	Sequence 181939,
27	340.5	6.5	882	15	US-10-334-143-9	Sequence 9, Appli
28	330.5	6.3	369	16	US-10-425-115-233034	Sequence 233034,
29	330.5	6.3	580	15	US-10-094-749-2012	Sequence 2012, Ap
30	323.5	6.1	726	15	US-10-467-434-17	Sequence 17, Appl
31	321	6.1	440	15	US-10-424-599-160260	Sequence 160260,
32	311	5.9	336	15	US-10-424-599-158918	Sequence 158918,
33	306	5.8	467	15	US-10-108-260A-3678	Sequence 3678, Ap
34	296	5.6	836	16	US-10-719-993-842	Sequence 842, App
35	296	5.6	836	16	US-10-719-993-843	Sequence 843, App
36	296	5.6	836	16	US-10-719-993-844	Sequence 844, App
37	295	5.6	856	16	US-10-719-993-845	Sequence 845, App
38	292.5	5.5	686	15	US-10-321-625-6	Sequence 6, Appli
39	279	5.3	718	13	US-10-087-192-333	Sequence 333, App
40	278	5.3	802	16	US-10-684-422-246	Sequence 246, App
41	271	5.1	1205	15	US-10-332-416-1	Sequence 1, Appli
42	267.5	5.1	814	17	US-10-781-581-185	Sequence 185, App
43	265	5.0	383	16	US-10-425-115-326274	Sequence 326274,
44	262.5	5.0	319	16	US-10-425-115-225159	Sequence 225159,
45	261.5	5.0	348	15	US-10-425-114-64367	Sequence 64367, A

#### ALIGNMENTS

RESULT 1  
US-10-210-281-74  
; Sequence 74, Application US/10210281  
; Publication No. US20040030096A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorman, Linda  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Zhong, Mei  
; APPLICANT: Patuturajan, Meera  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Pena, Carol E.A.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Sciore, Paul  
; APPLICANT: Stoney, David J.  
; APPLICANT: Taupier, Raymond J., Jr.  
; APPLICANT: Casman, Stacie  
; APPLICANT: Rothenberg, Mark E.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS  
; TITLE OF INVENTION: THE SAME  
; FILE REFERENCE: 21402-416D  
; CURRENT APPLICATION NUMBER: US/10/210,281  
; CURRENT FILING DATE: 2003-02-05  
; PRIOR APPLICATION NUMBER: 60/309,501  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 60/310,291  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 60/361,775

```
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/361,832
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,203
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,201
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,702
; PRIOR FILING DATE: 2001-08-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: CuraSeq1ist version 0.1
; SEQ ID NO 74
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-210-281-74

Query Match      57.0%; Score 3006; DB 15; Length 1132;
Best Local Similarity 55.3%; Pred. No. 6,8e-161;
Matches 626; Conservative 127; Mismatches 234; Indels 144; Gaps 22;

Qy 1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMVLKMKSVKA 60
Db 1 MPDQISVSEFIAETTEDYNSPTSTFTRLHNCRTNVTLLBEALQDRTALQKVKSVA 80

Qy 61 INSSGLAHVNEBEQYTOALEKFGNCVCRDDPDLGSAFLKFSVFTKELTALFKNLQNMN 120
Db 81 IYNSGQDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFVKFSTLTKESTLTKNLLQGLS 140

Qy 121 NIISFPLDSLKGLDKGVKGLDKKPFDKAWKDYETKITKEKEKEHAHLHGMIRTEISG 180
Db 141 HNVIFTLDSLKGDDUKGVKGLDKKPFDKAWKDYETKITKEKEKEHAHLHGMIRTEITG 200

Qy 181 AEIAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKAVES 240
Db 201 AEIAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKTADK 260

Qy 241 LKPSITETLSTDLHTIKQADERRLOIQRLILKSAQVQKE---DSQIRQSTAYSUHQ 297
Db 261 LKQVTEKLAADLYNIKQVDEEKKQLTALRDLIKSLQLDQKESRRDSQSRQG-GYSMHQ 319

Qy 298 PQGNKEHETERNGSLYKSDGIRKWQKRCVKNGLFTISHGTANRPPAKNLILLTCQVK 357
Db 320 LQGNKEYSEKGYLLKSDGIRKWQKRCVKNGLITISHATSNROPKAKNLILLTCQVK 379

Qy 358 TNPEKKCFDLISHDRTHYFAEDREQEQIWMVSLQNSKEEALNNAFPGDDNTGENNIVQ 417
Db 380 PNAEDKKSFDLISHNRTHYFAEDREQDYVAMISVLNTSKEEALTWAFRGEQSAGENSL-E 438

Qy 418 ELTKETIISEVQRMNTDNCDCGADPTWLTSTNLGILTCIECSGHRHRELVGHVSPMSQSLT 477
Db 439 DLTKAILEDVQRLPNDICCDGSEPTWLTSTNLGILTCIECSGHRHRELVGHVSIQSL 498

Qy 478 LDVLGTSSELLAKNIGNAGNEIMECCIPAEDSVKPNPNSDMNARKDYITAKYIERRYAR 537
Db 499 LDKLGTSELLAKNIGNNSFNDEIMANLPS-PSPKPTFSSDMTVRKEYITAKYVDHRFSR 557

Qy 538 KKHADNAAKLSLCEBAVTRDITFGLQAYADGVLDTEKIPIANGHEPDETALHLAVRSVD 597
Db 558 KTCSTSSAKNELLLEAIAKSRDOLLALIQYVAGVELWEPL-LEPGQELGETALHLAVRTAD 616

Qy 598 RTSHLIVFLVQNSGNLQKQGTGSHALHYCCLTQNAECLKLLLRGKASIEIANESGETP 657
Db 617 QTSHLVDFLVQNCNGLQKQGTALGNTVLHYCSMYSKPECLKLLLRKSKPTVDIVNAGETA 676

Qy 658 LDIAKRLKHECEELLTQALSRRFNHSHVVEYEWLHLLHEDLDESDDDDKDEKLPSENRE 717

; LDIAKRLKATQCEDLLSQAKSGKFNPHVHVEYEWLHLLHEDLDESDDDDKDEKLPSENRE 736
; DRPISFYQLGNSQLQSNVSLARDAANLAKEKQAFMPSILQNETYCALLSG---SPPPA 774
; PRQSFCHSSSISPD---KLALPGFSTPRDKQRL-----SYGAFTNQIFVSTSTD 784
; QPAAPSTTSAPPLPPRNVGK-----VQTASSANTLMKTNSVSDGGSRQSSSS 822
; SPTSP-TTEAPPLPPRNAGKGTGPPSTLPLSTQTSSTSGSSTLSKKRPPPPPPG-HKRTLS 842
; DPPAVHPLPLPLR-----VSTNTP-LTPTPPP 848
; DPPSPPLPHGPPENKAGVPGWNGDGGSSSKTTNKFEGLSQQSSSTSAKTALGPRVLPKLPQ 902
; PVA-KT-----PSVMEALSO-----PSKPAP-----PGISQIRPPP 878
; KVALRKTDHLSDKATIPPEIFQKSSQLAEPLQKPPPGDLPKPKFTELAPKFIQIGDLPPKP 962
; LPPQPP-SRLPOK-----KPAQCTD 897
; GELPPKQOLGDLPPKPOLSDLPKPKQMKDLPKPKQOLGDLAKSQTGDVSPKAAQPSVTL 1022
; KSTP--LTNKGQPRGPVDLSATEALGFLSNAMVLQPPAPMPRKSQATKLKPKRVKALYNC 955
; KSHPLDLSPNVQSRDAIQKQASEDSNDLTPTLP-ETVPPLPRKINTGKNKVRVKTIYDC 1081
; VADNPDELTSSEGDIIVDGEEDQEWIGHIDGDPGRKGAPFVSFVFIAD 1006
; QADNDELTTIEGEVIIVTGEEDQEWIGHIEGQPERKGVFPVSFVHLS 1132

RESULT 2
US-10-104-047-2951
; Sequence 2951, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2951
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-104-047-2951

Query Match      39.3%; Score 2071.5; DB 15; Length 903;
Best Local Similarity 43.4%; Pred. No. 2.7e-108;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;

Qy 1 MPDQISVSEFVAETHEDYKAPT-ASSTTRTAQCRNTVAABEALDVRMVLKMKSVK 59
Db 1 MPEQFSVAEFLVATAEALSSPAGAAFAAKPRYRGAALAREEILLEGDAILORIKKA 60

Qy 60 AINSSGLAHVNEBEQYTOALEKFGNCVCRDDPDLGSAFLKFSVFTKELTALFKNLQNM 119
Db 61 AIHSSGLAHVNEBEQYREAVESLGNHLSQNSHLSSTGFLNLAVFTREVAALFKNLQNL 120

Qy 120 NNIIISFPLDSLKGLDKGVKGLDKKPFDKAWKDYETKITKEKEKEHAHLHGMIRTEIS 179
Db 121 NNIVSFPLDSLKGLDKGVKGLDKKPFDKAWKDYETKITKEKEKEHAHLHGMIRTEIS 174

Qy 180 GAETAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKAVE 239
Db 175 PGEVAQDMQERRIFQLHMCPEYLLKAGESQKQGPDPFLQSLIKFFHAQHNFFQGWKAAQ 234

Qy 240 SLKPSIETLSTDLHTIKQADERRLOIQRLILKSAQVQKESQIRQST--AYSUHQ 297
```

```
Db 235 SUFFPIEKLAAVSHALHOAEDELQKLTQLRDSLRTGLQLESREBHLRKNSGCGYSIHQ 294
Qy 298 PQGNKEHGTERNGSLYKKSQDGIKRWKQKCSVQNGFLTISHGTANRPPAKNLNLTQCVK 357
Db 295 HQGNKQFGTEKVGFLYKKSQDGIKRWKQKCSVQNGFLTISHGTANRPPAKNLNLTQCVK 354
Qy 358 TNPEKKCFDLISHDRTHFOAEDBOEQIWMVSLQNSKEEALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTYHFOAEDBECEAWSVLQNSKDEALSAFLGEPGSGWS 414
Qy 410 TGENNIVOLTEKIISEVQRMGTNDVCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 469
Db 415 AGHDEPHDLTKLLIAEVKSPGNSQCCDCAADPTWLTSTNLGILTCIETCSGHIHRELGVH 474
Qy 470 YSPMQSLTDLVLTGSELLAKNIGNAGFNEIMECCLPADSVPKPNPNSDMARKDYITAK 529
Db 475 FSRMQSLTDLVLTGSELLAKNIGNAGFNEIMECCLPADSVPKPNPNSDMARKDYITAK 534
Qy 530 YIERYARKKHADNAKHLSCIAEVKTRDI FGLLOAYADGVDLTEKI PLANGHEPDETAL 589
Db 535 YVEHRFARR-----CTPEPQRLMTAICNRDLISVLEAFANGQDFGQPLPGPDAQAPEELVL 590
Qy 590 HLA VRSVDRTSLSHIVDFLVQNSGLDKOTGKSTALHYCCLTDNAECLKLLRKGASIEI 649
Db 591 HLA VQNAQSLPLVDFIIONGHLDKAAADGNTALHYAALYNQPDCLKLLKGRALVGT 650
Qy 650 ANESGETPLDIARLKHKECHELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKL 709
Db 651 VNEAGETALDIARLKHKECHELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKL 709
Qy 710 QPSNRREDRPIISFYQLGSNQLQSNVSLARDAANLAKEKQAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPAQAHWASGRGLDISNKTETVASL 739
Qy 769 GSPPPAQAPATTAPPLPRNVKQVTASSANTLWKTNSVSVGGSRQRSSDPPAVH 828
Db 740 GA-----ATPQGESEDCPPLPVKN-----SSRTLQ-----GCARHASGRDRSEVS 780
Qy 829 PPLPLRVLTSTNPLTTPPPVAKTPSVMEALSQPSKP--APPGISQ--IRPPLPPQP-- 883
Db 781 -----SLSSEAPETPESLGSPASSSLMSPL--EPGDFSQAPPNSEEGLREPFGTSRPSL 833
Qy 884 -----PSRLPQKKPAPGTDKSTPLTNKQPGFVDSATEALGPLSNAMVLQPPAPMPRK 938
Db 834 TSGTTPSEM--YLPVRFSSSESTRYRG-ARSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLPKRVKALNCVADNPDELTFSEGDIIVDGEEDQEWIHTGDPGRKGPV 998
Db 877 NVP-----VGITEGDSRGTSLPA 895
Qy 999 SPVHFIAD 1006
Db 896 SSVQLIQD 903
```

## RESULT 3

```
US-10-490-605-2
; Sequence 2, Application US/10490605
; Publication No. US20050019768A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCE, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; TITLE OF INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATL/SEN-A0121P-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; PRIOR FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USSN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-490-605-2
```

```
Query Match 39.3%; Score 2071.5; DB 17; Length 903;
Best Local Similarity 43.4%; Pred.No. 2.7e-108;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;
```

```
Qy 1 MPDQISVSEFAVETHEDYKAPT-ASSPTTRTAQCHNTVAALTEALDVRVLYKMKKSVK 59
Db 1 MPDQISVSEFAVETHEDYKAPT-ASSPTTRTAQCHNTVAALTEALDVRVLYKMKKSVK 60
Qy 60 AINSGLAHVNEEQYTQALBKFGCNCVCRDDPDLSGSAFLKPSVFTKELTALFKNLQNM 119
Db 61 AINSGLAHVNEEQYTQALBKFGCNCVCRDDPDLSGSAFLKPSVFTKELTALFKNLQNM 120
Qy 120 NNIISFPLDLSLLKGLDKGVKGDLLKPPFDKAWKDYETKITKEKEKHEKAKLHGMTRTEIS 179
Db 121 NNIISFPLDLSLLKGLDKGVKGDLLKPPFDKAWKDYETKITKEKEKHEKAKLHGMTRTEIS 174
Qy 180 GAETAEEMEKERRFPOLQMCVLLKVNIEIKKGVLDLLQNLIKYFHAQCNFFODGLKAVE 239
Db 175 PGEVAQDMQERRIFQLHMCBYLLKAGSQMKQGPDLQSLTKFPHAQNHFFODGKAAQ 234
Qy 240 SLKPSIETSLDHLTIKQADBEERRQLQLRDLKLSALQVEOKEDSQIRQST--AYSLLH 297
Db 235 SLFPIEKLAAVSHALHOAEDELQKLTQLRDSLRTGLQLESREBHLRKNSGCGYSIHQ 294
Qy 298 PQGNKEHGTERNGSLYKKSQDGIKRWKQKCSVQNGFLTISHGTANRPPAKNLNLTQCVK 357
Db 295 HQGNKQFGTEKVGFLYKKSQDGIKRWKQKCSVQNGFLTISHGTANRPPAKNLNLTQCVK 354
Qy 358 TNPEKKCFDLISHDRTHFOAEDBOEQIWMVSLQNSKEEALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTYHFOAEDBECEAWSVLQNSKDEALSAFLGEPGSGWS 414
Qy 410 TGENNIVOLTEKIISEVQRMGTNDVCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 469
Db 415 AGHDEPHDLTKLLIAEVKSPGNSQCCDCAADPTWLTSTNLGILTCIETCSGHIHRELGVH 474
Qy 470 YSPMQSLTDLVLTGSELLAKNIGNAGFNEIMECCLPADSVPKPNPNSDMARKDYITAK 529
Db 475 FSRMQSLTDLVLTGSELLAKNIGNAGFNEIMECCLPADSVPKPNPNSDMARKDYITAK 534
Qy 530 YIERYARKKHADNAKHLSCIAEVKTRDI FGLLOAYADGVDLTEKI PLANGHEPDETAL 589
Db 535 YVEHRFARR-----CTPEPQRLMTAICNRDLISVLEAFANGQDFGQPLPGPDAQAPEELVL 590
Qy 590 HLA VRSVDRTSLSHIVDFLVQNSGLDKOTGKSTALHYCCLTDNAECLKLLRKGASIEI 649
Db 591 HLA VQNAQSLPLVDFIIONGHLDKAAADGNTALHYAALYNQPDCLKLLKGRALVGT 650
Qy 650 ANESGETPLDIARLKHKECHELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKL 709
Db 651 VNEAGETALDIARLKHKECHELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKL 709
Qy 710 QPSNRREDRPIISFYQLGSNQLQSNVSLARDAANLAKEKQAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPAQAHWASGRGLDISNKTETVASL 739
Qy 769 GSPPPAQAPATTAPPLPRNVKQVTASSANTLWKTNSVSVGGSRQRSSDPPAVH 828
Db 740 GA-----ATPQGESEDCPPLPVKN-----SSRTLQ-----GCARHASGRDRSEVS 780
Qy 829 PPLPLRVLTSTNPLTTPPPVAKTPSVMEALSQPSKP--APPGISQ--IRPPLPPQP-- 883
Db 781 -----SLSSEAPETPESLGSPASSSLMSPL--EPGDFSQAPPNSEEGLREPFGTSRPSL 833
Qy 884 -----PSRLPQKKPAPGTDKSTPLTNKQPGFVDSATEALGPLSNAMVLQPPAPMPRK 938
Db 834 TSGTTPSEM--YLPVRFSSSESTRYRG-ARSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLPKRVKALNCVADNPDELTFSEGDIIVDGEEDQEWIHTGDPGRKGPV 998
Db 877 NVP-----VGITEGDSRGTSLPA 895
Qy 999 SPVHFIAD 1006
Db 896 SSVQLIQD 903
```



```
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 284
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2284

Query Match      20.4%; Score 1075.5; DB 15; Length 208;
Best Local Similarity 81.4%; Pred. No. 6.5e-53;
Matches 206; Conservative 1; Mismatches 1; Indels 45; Gaps 1;

Qy 754 MPSILQNETYTGALLSGSPPPAQAAPSTTSAPPLPPRVNGVKQTASSANTLWKTNSVSD 813
Db 1 MPSILQNETYTGALLSGSPPPAQAAPSTTSAPPLPPRVNGK----- 41

Qy 814 GGSQRSSDPPAVHPPPLPLRVSTNPLTTPPPPVAKTSPVMEALSQPSKPAPPGISQ 873
Db 42 -----DPLTTPPPPVAKTSPVMEALSQPSKPAPPGISQ 75

Qy 874 IRPPLPQPPSRPLPQKYPAPGTDKSTPLTNKGQPRGVDLSATEALGPLSNAMVLQPPA 933
Db 76 IRPPLPQPPSRPLPQKYPAPGADKSTPLTNKGQPRGVDLSATEALGPLSNAMVLQPPA 135

Qy 934 PMPRKSQATKLKPRKVKALYNCVADNPDELTFSEGDIIVDGEEDQEWIGHIDGDPGRK 993
Db 136 PMPRKSQATKLKPRKVKALYNCVADNPDELTFSEGDIIVDGEEDQEWIGHIDGDPGRK 195

Qy 994 GAFFVSVFHFAD 1006
Db 196 GAFFVSVFHFAD 208

RESULT 7
US-10-408-765A-2815
; Sequence 2815, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Sojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2815
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2815

Query Match      11.6%; Score 611.5; DB 16; Length 358;
Best Local Similarity 43.0%; Pred. No. 1.9e-26;
Matches 128; Conservative 17; Mismatches 18; Indels 135; Gaps 1;

Qy 68 HVNEEQYTALEKFGNVCRRDDPLGSAFLKFSVFTKEITALFPKNLI----- 116
Db 5 HVQNEENYAQLDKFGSNFLSRDNDLGTAFVKFSTLTKELSTLLKNLVRHFIFVRANAV 64

Qy 117 ----- 116
Db 65 TSAEQSNYGRKMIEDDPSTCKIFCLLHPRSLYPKATTPDLDCWEVLQPSLVLFSGIM 124
```

```
Qy 117 ----- 116
Db 125 LASTGKHGLVRGAWLDFRLPLYKPGVLEVPFVLCSEALLSEARKNEALLTDCLSASWLLQ 184

Qy 117 -----QNNNNIISFPLDSLLKGLGKVGKDLKKPDKAKWKDVEYTKITKIEKKEKHAHLHG 172
Db 185 QDRLOGLSHNVIFTLDSLKGLGKVGKDLKKPDKAKWKDVEYTKITKIEKKEKHAHLHG 244

Qy 173 MIRTEISGAETAEEMEKERRFFQLQMCCEYLLKVNNEIKIKKGVDDLQNLKIYFHAQCNF 230
Db 245 MIRTEITGAETAEEMEKERRLFLQMCCEYLLKVNNEIKIKKGVDDLQNLKIYFHAQCNW 302

RESULT 8
US-10-080-334-238
; Sequence 238, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Sureah G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
```



```
Query Match 10.6%; Score 558; DB 15; Length 764;
Best Local Similarity 23.2%; Pred. No. 5.2e-23;
Matches 174; Conservative 119; Mismatches 315; Indels 142; Gaps 15;

QY 26 FTTRTAQCRNTVAALIEALDVRMVLKXKSKVAINSSGLAHVENEEOYTQALEKFGN 85
Db 15 FRATIDEVETDVEIEAKLD-----KLKLCSGMVEAGKAYVSTSLFVSGVDRLSQQ 67

QY 86 CVCRRDPDGLSFAFLKFSVFTKELTALFKNLQIMNNIISFPDLSLLKGLKGVKGLDKRP 145
Db 68 --COGDTVISECLQRFADSLQEVVNMILFDQAQSVRQQLQSFVKEDVRKFK-ETKKQ 124

QY 146 PDKAWDYETKITKEKEKEHAKLHGIMRTISGAETAEEMEKERRFQOLQCMCYLLKV 205
Db 125 FDKVREDLELSLV-----RNAQPRHRPHEVE-----EATGALTLTRKCFRHLALDYVLQI 175

QY 206 NEIKIKGVLDLQNLKIFPHACNFFQDGLKAVESIKPSIETLSLDLHTIKQAQDEERQ 265
Db 176 NVLOAKKFEILDLSMFSFHAQSSFFQOQYSLHQDPMYKLAELDLQVLVDSAVEKRE 235

QY 266 LIQLRDILKALQVQKEDSQIRQSTAYSLHQPOGNKE-----HGTERNGSLYKSDGI 319
Db 236 M-----ERKHAALQORTLRDPSYDESKVEFDVDPASGVVMVEGLFKRASNA 281

QY 320 RKWOKRKSVKNGFLTISHGTANRPPAKL-NLLTCQVK--TNPEKKCFDLISHDRYH 376
Db 282 FKTNRWFISIQNSLVYQKLLKDALTVVVDLRLCSVKPCEDIERRFCEVLSPKSCM 341

QY 377 FOAEDQEQCIWMSVLQNSKEAL-----NNAFKDDNTGNNIVQ 417
Db 342 LQADSEKURQAWQAVQASIASAYRESPPDSYSEKLDRTASPTSSIDSATDTRERGK 401

QY 418 ELTKIIEVQRMGTNDVCCGAPDPTWLSNLGILTCIEGSHRELGVHYSPQSLT 477
Db 402 E---SVLQVQSVAGNSQCGDQDPDPRWASINLGVLLCIEGSHRSLGVHCSKVRSLT 458

QY 478 LVLGTSELLAKNTIGNAGFNEIMCCLPADSVKPNPQSDMNARKDYITAKYIERRYAR 537
Db 459 LDSWEPPELLKLMCELGSAVNQIYEAQCEGAGSRKPTASSSRQDKEAMIKDYVEKKFLR 518

QY 538 K----- 538
Db 539 KAPMAPALEAPRRVRVQKLRPHSSPRAPTARKVRLEPFLPCVAALSSEGAESSESGE 578

QY 539 KHADNAALKHSICE-----AVKTRDIFGLQAYADGVDLTEKIPLANGHEPDE 586
Db 579 ADGDTAEAWGLADVREHPGLLAHRAARDLPALAAALAHGAS-----VNWADAEDEGK 634

QY 587 TALHLAVRSVDRTSLSHIVDFVQNSGNLDKQTKGSTALHYCCLTDNAECLKLLRGKAS 646
Db 635 TPL---VQAVLGSLIVCFELQLQNGADVNQDRSDRGRAPLHATLLGRTGQVCLFLKRGAD 691

QY 647 IEIANESGETPLDIAKRLKHECEBLLTQA 676
Db 692 QHALDQEQRDPLAIAVQAANADIVTLRLA 721

RESULT 10
US-10-080-334-235
; Sequence 235, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkete, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, XiaoJia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
```

```
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Neha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 235
; LENGTH: 804
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-235
```

Query Match 10.4%; Score 551; DB 15; Length 804;

Best Local Similarity 22.5%; Pred. No. 1.4e-22;  
Matches 180; Conservative 119; Mismatches 323; Indels 178; Gaps 16;

```
QY 12 AETHEDYKAPTASSFTTRTAQCRNTVAALIEALDVRMVLKXKSKVAINSSGLAHVEN 71
Db 5 AEAAGDPSQP-RSCFRATIDEVETDVEIEAKLD-----KLKLCSGMVEAGKAYVST 56

QY 72 BEQVTOALEKFGGNCVCRDDPDGLSFAFLKFSVFTKELTALFKNLQIMNNIISFPDLSLL 131
Db 57 SRLFVSGVDRLSQQ--COGDTVISECLQRFADSLQEVVNMILFDQAQSVRQQLQSFV 114

QY 132 KGDLLKGVKGLDKKFPDKAWDYETKITKEKEKEHAKLHGIMRTISGAETAEEMEKER 191
Db 115 KEDVRKFK-ETKKQFDKVRDELELSLV-----RNAQPRHRPHEVE-----EATGALTLTR 164

QY 192 RFFQLQMCYEYLLKVNEIKIKGVLDLQNLKIFPHACNFFQDGLKAVESLKPSTETLSTD 251
```





339	TASPTSSIDSATDTRRGVKE-----SVLQRVQSVAGNSQCQDCGQDPDRWASINLGVLL	395
456	CIECSGIHRELGVHYSMPQSLTDLVLTSELLAKNIGNAGNEIMECCLPAAEDSVKPNP	515
396	CIECSGIHRLSGVHCSKVRSLTLDLDSWEPELLKLMCELGNSAVNQIYEAQCEGAGSRKPTA	455
516	GSDMNAWKDYTTAKYIERRYARK-----	538
456	SSSRQDEKAWIKDKYVENKFKLKPMPALPALEAPRRWRVQKCLRPHSSPRAPTARRKVRLE	515
539	-----KH	540
516	PVLPCVAALSSVGTLDLRKFRDRSLFCPPDELSLFSYFDAGAAGAPRKGAESESSGEAD	575
541	ADNAAKLHSLCE-----AVKTRDIFGLLQAVADGVLDLTEKIFLANGHEDEFTA	588
576	GDTEAEAWGLADVRELHFGLLAHRAARDLFPALAAALAHGAE-----VNWDAEDEGKTP	631
589	LHLAVRSYDRTSLHIVDFLVQNSGNLDKQTGKSTALHYCCITDNAECLKLLLRGRASTE	648
632	L---VQAVLGSGLIVCEFLQNGADVNRQDRSGRAPLHHATLLGRTGQVCLFKRGADQH	688
649	IANESGETPLDIAKRLKHEHCCELLTQA	676
689	ALDOEQRDPLAIVQAANADIVTLLURLA	716

RESULT 12

US-10-1776-306-11

Sequence 11, Application US/10176306

Publication No. US20030130485A1

GENERAL INFORMATION:

APPLICANT: Meyers, Rachel A.

APPLICANT: Curtis, Rory A. J.

APPLICANT: Glucksmann, Maria Alexandra

APPLICANT: Bandaru, Rajasekhhar

APPLICANT: Kapeller-Libermann, Rosana

TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF

FILE REFERENCE: 10448-195001

CURRENT APPLICATION NUMBER: US/10/176,306

CURRENT FILING DATE: 2002-06-20

PRIOR APPLICATION NUMBER: 10/001,137

PRIOR FILING DATE: 2001-11-14

PRIOR APPLICATION NUMBER: PCT/US01/45291

PRIOR FILING DATE: 2001-11-14

PRIOR APPLICATION NUMBER: 60/248,362

PRIOR FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: 60/248,331

PRIOR FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: 60/248,365

PRIOR FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: 60/250,077

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 60/250,327

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 60/250,176

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 10/023,617

PRIOR FILING DATE: 2001-12-18

PRIOR APPLICATION NUMBER: PCT/US01/49416

PRIOR FILING DATE: 2001-12-18

PRIOR APPLICATION NUMBER: 60/256,249

PRIOR FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: 60/256,405

PRIOR FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: 10/083,248

PRIOR FILING DATE: 2001-10-22

PRIOR APPLICATION NUMBER: PCT/US01/46717

PRIOR FILING DATE: 2001-10-22

PRIOR APPLICATION NUMBER: 60/242,324

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/242,518

PRIOR FILING DATE: 2000-10-23

```

; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 834
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-306-11

Query Match      10.0%; Score 529; DB 14; Length 834;
Best Local Similarity 21.4%; Pred. No. 2.5e-21;
Matches 175; Conservative 119; Mismatches 315; Indels 210; Gaps 15

Qy 26 FTTTACQNTVAAIEBALDVRMVLVYMKKSKVAINSSGLAHVENEBEQYTOALEKFCGN 85
Db 15 FRATIDEVETDVBEIAKLD-----KLVKLCSGMVEAGKAYVTSRSLFVSGVRDLSSQ 67
Qy 86 CVCRRDDPDLGSFAFLKFSVFTKELTALFKNLIQNNNNIISFPLDLSLLKGLKGVGDLKKP 145
Db 68 --CGDVTIVISECLQRFADSLQEVVNYHMLFPDQAQSRVROQLQSFVKEDVRKF-ETKKQ 124
Qy 146 FDKAWKDYEYTKITKIEKEKEHAKLHGMIRTEISGAETAEEMEKERRFFQLQCMCEYLLKV 205
Db 125 FDKVREDELSLV-----RNAQAPRHPHPEVE---EATGALTLTRKCFRHLADYVLQI 175
Qy 206 NEIKIKGVLLQNLIKYFHAQCFFQDGLKAVESLKPSIETLTDL-----HTIKQAQ 259
Db 176 NVLQAKKKFILLDSMLPFMAHQSSFFQOQYSLHLQDLPYMKKLAELDQLVIDSAVEKRE 235
Qy 260 DEERRQLIQRLDKLSALQVEQKEDSIOIRSTAYSLHQPQNKKEHGTERNGSLYKKSQGI 319
Db 236 MERKHAALQORTLLQDFSYDESKVEPDV-----DASGVVMWEGYLFKASNA 282
Qy 320 RKVQKPKCSYKNGFLTISHGTANRPAPKL-NLLTCQVK--TNPEEKKCFDLISHDRTYH 376
Db 283 FKTWNRRWFSTQNSQLVYQKKLDKALTVMVDDLRALCSVKPCEDIERRCFEVLSPTKSCM 342
Qy 377 FOADEQECCIWMVSLQNSKEAL-----NNAFKGGDDNTGNNIVQ 417
Db 343 LQADSEKLQAWQVQVQASIASVRSBSCYSERLORTASPSTSDSATDTRRGVKG 402
Qy 418 ELTKETISEVQRMGTNDVCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSPMQSLT 477
Db 403 E---SVLQVQSVAGNSQCQDQCPDPRWASINLGVLLCIECSGIHRSGLGVHCSKVRSLT 459
Qy 478 LDVLGTSLELLAKNIGNAGNEIMECCLPABEDSVKPNPGSDMNARKOYITAKYIERRYAR 537
Db 460 LDSWEPPELLKMLCGLNSAVNQIYEAQCEGASRKPTASSRQDKBAWIKDKYVEKFKLR 519
Qy 538 K----- 538
Db 520 KAPMAPALEAPRRWRVQKCLRPHSSPRAPTARKVRLEPVLPCVAALSSVGTLOKRPDR 579
Qy 539 ----- 538
Db 580 SLFCPEDELDFSYFDGAAAGAGAPRSLSDSLGSGSDGSDVLAFGSGSVVDVSTEBEG 639
Qy 539 -----KHANNAKHLSLCE-----AVKTRDIFCLLQAYADGVDLTKEIP 577
Db 640 ASESSESGADGDTAEAWGLADVRELHPGLLAHRAARARDLPALAAALAHGAE-----VN 695
Qy 578 LANGHEPDETALHLAVRSVDRTSLIHIVDFLVQNSGNIJDQTKGKSTALHYCCLTDNAECL 637
Db 696 WADADEGKTPL---VQAVLGSSLIYCEFFLLQNGADVNRQDSRGRAPLHATLLGRTGOV 752
Qy 638 KLLIRGKASTEIANESGETPLDTAKRLKHEHCBEELLTQA 676
Db 753 CLFLKRGADQHALDQEQRDPLATAVQAANADIVTLLRLA 791

```

RESULT 13  
US-10-080-334-237

; Sequence 237, Application US/10080334  
 ; Publication No. US20040002584A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pena, Carol E. A.  
 ; APPLICANT: Shimkets, Richard A.  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Shenoy, Suresh G  
 ; APPLICANT: Kekuda, Ramesh  
 ; APPLICANT: Spytek, Kimberly A.  
 ; APPLICANT: Vernet, Corine A. M.  
 ; APPLICANT: Malyankar, Uriel M.  
 ; APPLICANT: Guo, Xiaojia  
 ; APPLICANT: Gusev, Vladimir Y  
 ; APPLICANT: Casman, Stacie J  
 ; APPLICANT: Boldog, Ferenc L  
 ; APPLICANT: Furtak, Katarzyna  
 ; APPLICANT: Tchernev, Velizar T  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Gangolli, Esha A  
 ; APPLICANT: Padigaru, Muralidhara  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Baumgartner, Jason C.  
 ; APPLICANT: Gerlach, Valerie  
 ; APPLICANT: Spaderna, Steven K  
 ; APPLICANT: Zerhusen, Bryan D  
 ; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
 ; FILE REFERENCE: Using the Same  
 ; FILE REFERENCE: 21402-275  
 ; CURRENT APPLICATION NUMBER: US/10/080,334  
 ; PRIOR FILING DATE: 2002-02-21  
 ; PRIOR APPLICATION NUMBER: 60/270,523  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: 60/322,712  
 ; PRIOR FILING DATE: 2001-09-17  
 ; PRIOR APPLICATION NUMBER: 60/311,980  
 ; PRIOR FILING DATE: 2001-08-13  
 ; PRIOR APPLICATION NUMBER: 60/330,307  
 ; PRIOR FILING DATE: 2001-10-18  
 ; PRIOR APPLICATION NUMBER: 60/278,796  
 ; PRIOR FILING DATE: 2001-03-26  
 ; PRIOR APPLICATION NUMBER: 60/281,521  
 ; PRIOR FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/276,677  
 ; PRIOR FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/311,595  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/270,220  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: 60/274,295  
 ; PRIOR FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: 60/318,526  
 ; PRIOR FILING DATE: 2001-09-10  
 ; PRIOR APPLICATION NUMBER: 60/286,548  
 ; PRIOR FILING DATE: 2001-04-25  
 ; PRIOR APPLICATION NUMBER: 60/291,765  
 ; PRIOR FILING DATE: 2001-05-17  
 ; PRIOR APPLICATION NUMBER: 60/270,797  
 ; PRIOR FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: 60/276,400  
 ; PRIOR FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/270,810  
 ; PRIOR FILING DATE: 2001-02-23  
 ; NUMBER OF SEQ ID NOS: 388  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 237  
 ; LENGTH: 834  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-080-334-237

Query Match 10.0%; Score 529; DB 15; Length 834;  
 Best Local Similarity 21.4%; Pred. No. 2.5e-21;  
 Matches 175; Conservative 119; Mismatches 315; Indels 210; Gaps 15;

QY 26 FTTRTAOCRNTVAABEALDVRMLVKMKSVKAINSSGLAHVNEBEQYTOALEKEGCGN 85  
 DB 15 PRATIDEVETDVVEIEAKLD-----KLVKCSGMEVAGKAYVSTSRFLVSGVGRDLSQQ 67  
 QY 86 CVCRRDDPDLSAFKLSVFTKELTALPKNLQNNNIISFPLDLSLLGDKLKGVDLKKP 145  
 DB 68 --CQGDVISECLQRFADSLQEVVNYHMLFDQAQSRVROQLQSFVKEDVRKFK-ETKKQ 124  
 QY 146 PDKAWDYETKITKIEKEKEHAKLHGMIRTEISGASIAEEMEXERFFQLOMCEYLLKV 205  
 DB 125 FDKVREDLELSLV-----RNAQAPRHPHEVE-----EATGALTLTRKCFRHLADLYVLIQI 175  
 QY 206 NEIKIKKGVLLQNLKLYFYHAQCNFFODGLKAVESLKPSTETLTDL-----HTIKQAO 259  
 DB 176 NVLQAKKKFEILDLSMLSFMAQSSFFQGGYSLHLQLDPYMKKLAEBLDQLVIDSAVEKRE 235  
 QY 260 DEERRQLIQLRDILKSALQVEKEDSROIROSTAYSLHQPOGNKEHGHTERNGSLYKSDGI 319  
 DB 236 MERKHAALQORTLLQDFSYDESKVEFDV-----DAPSGVVMGYLFRASNA 282  
 QY 320 RKVWQKRCVNGFLTISHGTANRPPAKL-NLLTCQVK--TNPEKKCPDLISHDRYH 376  
 DB 283 FKTWNRWFSTQNSQLVYQKKLDALTVMVDDLRLCSVKPCEDIERRFCEVLSPPTKSCM 342  
 QY 377 FQAEDEQECQIIMSVLQNSKEAL-----NNAFKGGDDNTGNNIVQ 417  
 DB 343 LQADSEKLQAWQAVQASIASAYRESPDCYSERLDRTPASTSTSSIDSATDTRGVKG 402  
 QY 418 ELTKEIIEVQRTMGNDVCCDGPDTWLSTNLGILTCIECSGIRHRELGVHYSPMOSLT 477  
 DB 403 E---SVLQRVQSVAGNSQCGDQGPDRWASINLGVLLCTIECSGIRHSLGVHCVSKVRSLT 459  
 QY 478 LDVLGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNAKDYITAKYIERRYAR 537  
 DB 460 LDSWEPELLKMLCGLNSAVNQIYEAQCEGAGSRKPTASSSRQDKBAWIKDKYVEKKPLR 519  
 QY 538 K-----KADNAAKLHSLCE-----AVKTRDIFGLLQAYADGVLTETKIP 577  
 DB 640 AESSESSGEADGDTAEAWGLADYRELHPGLLAHRAARADLPALAAALAHGAE-----VN 695  
 QY 578 LANGHEPDETALHLAVRSVDRSTSLHIVDFLVQNSGNDLKOTGKGSTALHYCCLTDNABCL 637  
 DB 696 WADAEDEGKTPL---VQAVLGGSLIVCEFLQNGADVNRQDSRGRAPLHATLLGRGTQV 752  
 QY 638 KLLRGKASIBIANESGETPLDIARLKHCEHELLTOA 676  
 DB 753 CLFLKRGADQHALQEQDRDPLAIAVQAANADIVTLRLA 791

RESULT 14  
 US-10-467-434-4  
 ; Sequence 4, Application US/10467434  
 ; Publication No. US20040092715A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, Li; WARREN, Bridget A.;  
 ; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom;  
 ; APPLICANT: YUE, Henry; BURFORD, Neil;  
 ; APPLICANT: LEE, Sally; RICHARDSON, Thomas W.;  
 ; APPLICANT: LAL, Preeti G.; NGUYEN, Dannie B.;  
 ; APPLICANT: YANG, Junning; HAFALIA, April J.A.;  
 ; APPLICANT: YONG, Craig H.; GURURAJAN, Rajagopal;  
 ; APPLICANT: BAUGHN, Mariah R.; WANG, Yumei E.;  
 ; APPLICANT: YAO, Monique G.; THANGAVELU, Kavitha;  
 ; APPLICANT: SWARNAKAR, Anita; GRIFFIN, Jennifer A.;

APPLICANT: FORSYTHE, Ian J.; EMERLING, Brooke M.;  
APPLICANT: CHAWLA, Narinder K.;  
TITLE OF INVENTION: Intracellular Signaling Molecules  
FILE REFERENCE: PF-0897 USN

CURRENT APPLICATION NUMBER: US/10/467,434

CURRENT FILING DATE: 2003-08-06

PRIOR APPLICATION NUMBER: PCT/US02/03966

PRIOR FILING DATE: 2002-02-07

PRIOR APPLICATION NUMBER: US 60/267,925

PRIOR FILING DATE: 2001-02-08

PRIOR APPLICATION NUMBER: US 60/274,435

PRIOR FILING DATE: 2001-03-09

PRIOR APPLICATION NUMBER: US 60/277,819

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: US 60/281,326

PRIOR FILING DATE: 2001-04-03

PRIOR APPLICATION NUMBER: US 60/291,195

PRIOR FILING DATE: 2001-05-15

PRIOR APPLICATION NUMBER: US 60/291,550

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: US 60/293,591

PRIOR FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: US 60/295,348

PRIOR FILING DATE: 2001-06-01

NUMBER OF SEQ ID NOS: 36

SOFTWARE: PERL Program

SEQ ID NO 4

LENGTH: 834

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No: 6892116CD1

US-10-467-434-4

Query Match 10.0%; Score 529; DB 15; Length 834;

Best Local Similarity 21.4%; Pred. No. 2.5e-21;

Matches 175; Conservative 119; Mismatches 315; Indels 210; Gaps 15;

26 FTTRTAQCRNTVAATEALDVRMVLVYKMKSVKAINSSGLAHVNEBQYTOALEKFCGN 85

15 FRATIDEVETDVEAEKLD-----KLVKLCGMEAGKAYVSTSLFVSGVNDLSQQ 67

86 CVCDDPDLGSAFLKFSVFTKELTALFKNLQNMNNIISFPLDSLKGLDKGVKGLKPP 145

68 --CGQDTVISECLQFADSLQEVVNYHMLFDQAQSVRQQLQSPVKEDVRKEK-ETKKQ 124

146 FDKAWKDYETIKIEKKEKHAHGMIRTEISGAETAEWEKERRFFQLOMCEYLLKV 205

125 FDKVREDLESLV-----RNAQAPRRPHEVE---EATGALTLTRKCFRHLADYVLQI 175

206 NEIKIKGVLDLQNLIKYFHAQCNPFDQGLKAVESLKSITLSTDL-----HTIKQAQ 259

176 NVLQAKKFEILDMSLWFAHQSSFFQGYSLHLQDPMYKLAELDLQVLIDSAVEKRE 235

260 DEERQLQLRDLILKALQVEKDSQIRQSTAYSLHQPQNGKHGTERNGSLYKKS DGI 319

236 MERKHAALQORTLLQDPSYDESKVEFDV-----DAPGGVMEGYLFKASNA 282

320 RKVQKRCVKNQFLTISHGTANRPPAKL-NLITCQVK--TNPEEKCFDLISHDRYH 376

283 FKTWNRWFSIQNSLVYQKLLKDALTVVDDLRCLSVKPCEDIERRFCFVLSPTKSCM 342

377 FOADEQECQIWMVSLONSKEAL-----NNAPKGDNDTENNIVQ 417

343 LOADEKLRQAWQVQVQASIASAVNESPDSCYSERLDRTAGSPSTSSIDATDTRERGK 402

418 ELTKEIIEVQRMGTNDVCCGAPDPPTWLSTNLGILTCIECSGTHRELGVHYSPMSLT 477

403 E---SVLQVQSVAGNSQCGCGQDPDPRWASINLVLLCIECSGTHRELGVHYSPMSLT 459

478 LDVLGTSSELLAKQIGNAGFNEIMECCPLPAEDSVKPNFGSDMNAKQYITAKYIERRYAR 537

Db 460 LDSHEPELLKLMCELGNSAVNQIYEAQCEGAGSRKPTASSSRQDKBAWIKDYVEKKFLR 519

Qy 538 K----- 538

Db 520 KAPWAPALEAPRRWRVQKCLAPHSSPRAPTARRKVRLEPVLPCVAALSSVGTLDKRFRRD 579

Qy 539 ----- 538

Db 580 SLFCPDELDSLFSYFDAGAAGAPRSLSSDSGLGSSDGLAFSGSVVDSVTEBEG 639

Qy 539 -----KHADNAAKLHSLCE-----AVKTRDIFGLQAYADGVDLTEKIP 577

Db 640 ABESESSGEADGTEAEAWGLADVRELHPGLLAHRAARADLPALAAALAHGAE-----VN 695

Qy 578 LANGHEPDETALHLAVRSVDRSTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECL 637

Db 696 WADADEGKTP-----VQAVLGSLIVCEFLQNGADVNRDSRGRAPLHATLLGRTQV 752

Qy 638 KLLLRGKASIEIANESGETPLDIAKRLKHEHCEBELLTOA 676

Db 753 CLFLKRGADQHALDQEQRDPLAIAVQAANADIVTLRLA 791

#### RESULT 15

US-10-739-930-6184

Sequence 6184, Application US/10739930

Publication No. US20040216190A1

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT

FILE REFERENCE: 38-21(53377)B

CURRENT APPLICATION NUMBER: US/10/739,930

CURRENT FILING DATE: 2003-12-18

NUMBER OF SEQ ID NOS: 11088

SEQ ID NO 6184

LENGTH: 775

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

OTHER INFORMATION: Clone ID: ARATH-23APR03-C24093\_1.p

US-10-739-930-6184

Query Match

Best Local Similarity 9.7%; Score 509.5; DB 16; Length 775;

Matches 190; Conservative 123; Mismatches 292; Indels 177; Gaps 24;

Qy 37 VAAITEALDVRMVLVYKMKSVKAINSSGLAHVNEBQYTOALEKFCGNVCYCRDDPDL-- 94

Db 17 VCSLEGTADLKDRCQKLYKGVKFMGTGLGEASKGESAFACLEBEFGGG--PDDPISLS 73

Qy 95 --GSAPFLKFSVFTKELTALFKNLQNMNNIISFPLDSLKGLDKGVKGLKPPDKAWKD 152

Db 74 IGGPVIKFINAURELASYKEFLCSQVHVLLELMNFIISVDLQSEAK-ESRHRFDKAAHS 132

Qy 153 YETKITIKIEKKEKHAHGMIRTEISGAETAEWEKERRFFQLOMCEYLLKVAEIKKK 212

Db 133 YDQSRKFSVLSKN-----TRGEIV-ABLEEDLENSKSTFEKSRNLVNSLTIEAKK 184

Qy 213 GVDLLQNLIKYFHAQCNPFDQGLKAVESLKSITLSTDLHTIKQ-----AQDEERRQLIQ 268

Db 185 KYEPLESISAIMDAHLRVFKLYGVDLLQLEPFIHQILTYAQSKESQSKIEQDLRLARIQE 244

Qy 269 LRDILKALQVEKDSQIRQSTAYSLHQPQNGKHG-----TE-----RNGSL 312

Db 245 FR-----TQSELDLSQQLVANAES--SGANGNRVGNIPYKNTETSLTADKEVIKQYL 295

Qy 313 YKXSDGIRKWKQKCKSVKNGKFLTISHGTA-----FVLDSDHSGMYVYRTNKGMSHHHSGSDHNTGVGR 342

Db 296 LKRSSSURTQW-KKKF-----FVLDSDHSGMYVYRTNKGMSHHHSGSDHNTGVGR 349

Qy 343 -----NRPPA-----KLNLLTCQVKTNPBE---KCCFDLISHDRTYHFOADEQEC 385

Db 350 FRARHNRSGSLTEGSLGYNTIDLRSLIKLDAEDMDLRLCFRIISPKQTYTLQAENGADR 409  
Qy 386 QIWNQSVLQNSKEEALN-----NAPKGDNDTGENNIVQELTKEI 423  
Db 410 MDWVNKITKAIGTLNLSHFLQOSPVRYLKONSSAPANAVVSGQIRHNDNRQNIIGDDV 469  
Qy 424 ISEVQRTGNVCCDCGAPDPTWLSNLTGILTCIECSGIHRELGVHYSQMOSLTLD--VL 481  
Db 470 STILRGLFGNNAECAECNAPEDWASLNLGVLLCQCSCGVHRNLGVHISKVRSLSLDVKW 529  
Qy 482 GTSELLLLAKNIGNAGFNEIMECCLPAEDSV-----KPNPGSDMMARKDYITAK 529  
Db 530 EPTILDLPRLNGVVCNSLWEGLLHLDCCEDGSALSASVSKPCPEDSFSVKEKYILGK 589  
Qy 530 YIERRYARKHAD-NAAKHLSLCEAVKT---RDIPLLOQAYAD----- 568  
Db 590 YLEKALVIKDESEANLSAASRIWEAVQSRNIREIYRLIVTTGVDVNIINTKFDIDTIDAY 649  
Qy 569 -GVDLTEKIPLANGHEPDE-----TALHLAVRSVDRSTLSLHIVDFLVQN 610  
Db 650 HHIDAEEK-AVKRHDPTVCORIKESNEPRSCLOQCSLLHVACHIGDSV---LLELLLOP 705  
Qy 611 SGNLDKQTKGSTALHYCLTDNAECLKLLLRGKASIEIANESGETPLDIAKRLKHECE 670  
Db 706 GADLNIRDYHGRTPLHHCISSGNHKFAKILLRRGARPSTIEDDGGUSVLERAMEMGAITDE 765  
Qy 671 EL 672  
Db 766 EL 767

Search completed: August 4, 2005, 08:34:44  
Job time : 108 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 16:31:46 ; Search time 7497 Seconds  
(without alignments)  
5107.730 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 1006  
Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAPFVSVFHTAD 1006

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
  
Searched: 34239544 seqs, 19032134700 residues  
Word size: 1

Total number of hits satisfying chosen parameters: 68477535

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool\_h/US0914042/runat\_04082005\_090747\_2671/app\_query.fasta\_1.1159  
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0914042 @CGN 1 1 4326 @runat\_04082005\_090747\_2671 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsel1:\*  
9: gb\_gsel2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	202	20.1	1167	5	BU508193
2	198	19.7	901	5	BU172807
3	177	17.6	556	1	AL044307
4	165	16.4	846	5	BU155022
5	160	15.9	689	5	BU703510
6	160	15.9	731	7	CN538273
7	160	15.9	773	7	CN460643
8	160	15.9	781	7	CO429713
9	157	15.6	582	5	BP280849

LOCUS	BU508193	AGENCOURT_10128476	NIH_MGC_71	Homo sapiens	cdna	clone	EST 12-SEP-2002
DEFINITION	5', mRNA sequence.						
ACCESSION	BU508193						
VERSION	BU508193.1	GI:22814426					
KEYWORDS	EST.						
SOURCE	Homo sapiens						
ORGANISM	Homo sapiens (human)						
REFERENCE	1 (bases 1 to 1167)						
AUTHORS	NIH-MGC http://mgs.nci.nih.gov/.						
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)						
JOURNAL	Unpublished (1999)						
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLNL4057 row: p column: 08 High quality sequence stop: 762.						
FEATURES	Location/Qualifiers						
source	1..1167						
	/organism="Homo sapiens"						
	/mol_type="mRNA"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:6502231"						

#### ALIGNMENTS

LOCUS	BU508193	AGENCOURT_10128476	NIH_MGC_71	Homo sapiens	cdna	clone	EST 12-SEP-2002
DEFINITION	5', mRNA sequence.						
ACCESSION	BU508193						
VERSION	BU508193.1	GI:22814426					
KEYWORDS	EST.						
SOURCE	Homo sapiens						
ORGANISM	Homo sapiens (human)						
REFERENCE	1 (bases 1 to 1167)						
AUTHORS	NIH-MGC http://mgs.nci.nih.gov/.						
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)						
JOURNAL	Unpublished (1999)						
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLNL4057 row: p column: 08 High quality sequence stop: 762.						
FEATURES	Location/Qualifiers						
source	1..1167						
	/organism="Homo sapiens"						
	/mol_type="mRNA"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:6502231"						

```

/tissue type="leiomyosarcoma"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 7,39e-201 Length: 1167
Score: 202.00 Matches: 263
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 20.08% Indels: 2
DB: 5 Gaps: 0

```

US-09-914-042-1 (1-1006) x BU508193 (1-1167)

```

Qy 411 GlyGluAenAenIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMet 430
Db 24 GGAGAAATAACATCGTCCAGAACTGACAAAGGAGATCATCTCAGAAAGTGACAGAGATG 83
Qy 431 ThrGlyAenAenValCysCysAspCysGlyAlaProAspProThrTriPLeuSerThrAen 450
Db 84 ACGGCATAGACGCTGCTGACGTGTGGGGCCAGATCTTACATGCTTCCACCAAC 143
Qy 451 LeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyr 470
Db 144 CTGGCATCTCTGACCTGATCGAGTGTCCGGAATCCACCGAGAGCTGGGGTTCATTAT 203
Qy 471 SerPro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLy 490
Db 204 TCCAG-GATGAGTCCCTGACCTTAGATGTACTGGGAACATCTGAGCTGCTGCCAA 262
Qy 490 sAenIleGlyAenAlaGlyPheAenGluIleMetGluCysCysLeuProAlaGluAspSe 510
Db 263 GAATATTGGGAATGAGGCTTTAATGAGATCATGGAATGTTGCTTACCAGCTGAGGACTC 322
Qy 510 rValysProAenProGlySerAspMetAenAlaArgLysAspTyrIleThrAlaLysTyr 530
Db 323 AGTCAAAACCCACCCAGGAGCGACATGAATGTCAGAAAGGACTACATCACAGCCCAAGTA 382
Qy 530 rIleGluArgArgTyrAlaArgLysLysHisAlaAspAenAlaLysLeuHisSerLe 550
Db 383 CATCGAGAGGAGATACGCAAGGAAGAACCGGATACCGGCGGAAGCTTTCACAGTCT 442
Qy 550 uCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyVa 570
Db 443 TTGCGAGCCCGTCAAAACAGAGATATTTTGGATTGCTCCAAAGCTTATGCTGATGGTGT 502
Qy 570 lAspLeuThrGluLysIleProLeuAlaAenGlyHisGluProAspGluThrAlaLeuHi 590
Db 503 GGATCTTACGGAAAAAATCCCACTGGCCAAACGACATGACCGGATGAACCGGCCCTCCA 562
Qy 590 sLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAs 610
Db 563 CTTGTCAGTCAGATCCGCGGATCGAAGCTCTCTTACATGTPAGACTTTTGTAGTTTCAGAA 622
Qy 610 nSerGlyAenLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLe 630
Db 623 CAGTGGGAACCTGGATAACACAGAGGAAAGGACAGACAGCCCTGCCTACTGCTGCT 682
Qy 630 uThrAspAenAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAl 650
Db 683 GACCACAAATGCCAGTGCTCAAGTGTCTCTGCGGGGAAGGCTCCATCGAGATAGC 742
Qy 650 aAenGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysG1 670
Db 743 AAACAGTTCAGAGAGACTCCCTGTCGACATTCGCAAGCGCTTCAAGCAGCAGCACTGTGA 802
Qy 670 uGluLeuLeuThr 674
Db 803 GGAGCTGCTGACC 815

```

## RESULT 2

## LOCUS

```

DEFINITION BUI72807 901 bp mRNA linear EST 04-SEP-2002
AGENCOCURT 7966859 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6170038
5', mRNA sequence.

```

## ACCESSION

BUI72807.1 GI:22686791

## VERSION

EST.

## KEYWORDS

SOURCE

## ORGANISM

Homo sapiens (human)

## REFERENCE

1 (bases 1 to 901)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapba-remail.nih.gov](mailto:cgapba-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LUAMJ3536 row: b column: 23

High quality sequence stop: 609.

## FEATURES

## source

```

1..901
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6170038"
/tissue_type="retinoblastoma"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 9,31e-197 Length: 901
Score: 198.00 Matches: 232
Percent Similarity: 99.15% Conservative: 0
Best Local Similarity: 99.15% Mismatches: 1
Query Match: 19.68% Indels: 2
DB: 5 Gaps: 0

```

US-09-914-042-1 (1-1006) x BUI72807 (1-901)

```

Qy 123 lIeSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGlyAspLeu 142
Db 14 ATCTCTCTCCCTTTGGACAGTTTCTGAGGGGACCTGAAAGGAGTGAAGGGGATCTG 73
Qy 143 LysLysProPheAspLysAlaTriPLeuAspTyrGluThrLysIleThrLysIleGluLys 162
Db 74 AAAAAGCCTTTTGATAAAGCTTGAAGGACTATGAAACAAAATAACCAAGATAGAAAG 133
Qy 163 GluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGlu 182
Db 134 GAGAAAAGGAAACACGCCAAGCTCATGGGATGATTCGAGCTGAAATTAAGCGGACGGAA 193
Qy 183 lIeAlaGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGluTyrLeu 202
Db 194 ATTCCGAGAGATGGAAAGAGAGAGGCGCTTCTTCAGACTACAGATGTGCGAGTATCTG 253
Qy 203 LeuLysValAenGluIleLysIleLysGlyValAspLeuLeuGlnAenLeuLys 222
Db 254 CTGAAGGTCAACGAAATCAAGATTAAGGAGGAGTAGATTTACTTTCAGATCTGTATCAA 313

```

QY 223 TyrPheHisAlaGlnCysAsnPhenGlnAspGlyLeuLysAlaValGluSerLeuLys 242  
 Db 314 TACTTTTCATGCCCAATGCAATTTTTTTTCAGGATGGACTCAAAAGCGGTGGAAGCCTCAAA 373  
 QY 243 ProSerileGluThrLeuSerThrAspLeuHisThrileLysGlnAlaGlnAspGluGlu 262  
 Db 374 CTTTCATTGAACCGCTGCTACGGATCTTACAGGATCAACAGGCCGCCAGATGAAGAA 433  
 QY 263 ArgArgGlnLeuileGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLys 282  
 Db 434 AGAAGGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGATTCGCGATTGGAACAGAA 493  
 QY 283 GluAspSerGlnIleArgGlnSerThrAlaTySerLeuHisGlnProGlnGlyAsnLys 302  
 Db 494 GAGGACTCCCAAAATTCGTCAGAGCACAGCTTATAGCTTATACATGACGCTTCAGGGAACAAG 553  
 QY 303 GluHisGlyThrGluAtGAspGlySerLeuTyLysLysSerAspGlyLeuArgLys-Va 322  
 Db 554 GAACATGGGACCGAGCGAAGCGAGCTCTACAGAAGAGTGACGGATCCGAAN-AGT 612  
 QY 322 lTTPGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAl 342  
 Db 613 GTGCAGAAAAGGAAATGTTTCAGTTAAATGTTTCTGACCATATATCCCATGTTACCGC 672  
 QY 342 aAsnArgProAlaLysLeuAsnLeuLeuThrCysGln 355  
 Db 673 TAACCGGCTCTCGAAGCTCAACCTGCTAACCTGCCAG 712

RESULT 3  
 LOCUS AL044307 556 bp mRNA linear EST 04-SEP-2003  
 DEFINITION DKFZp343402129 r1 434 (synonym: htes3) Homo sapiens cDNA clone  
 DKFZp343402129 5', mRNA sequence.  
 ACCESSION AL044307  
 VERSION AL044307.1 GI:5432530  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 556)  
 AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
 TITLE EST (Koehrer, et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: MIPS  
 MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by BPFZ (Biomedical Research Center at the Charite,  
 Berlin/Germany) within the cDNA sequencing consortium of the German  
 Genome Project.  
 No sl sequence available.  
 This clone (DKFZp343402129) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
 source  
 1..556  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp343402129"  
 /tissue\_type="testis"  
 /dev\_stages="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="434 (synonym: htes3)"  
 /note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

Alignment Scores:  
 Pred. No.: 7.25e-175  
 Score: 177.00 Matches: 556 177

ORIGIN

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 17.59% Indels: 0  
 Db: 1 Gaps: 0  
 US-09-914-042-1 (1-1006) x AL044307 (1-556)  
 QY 166 GluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGlyAlaGluIleAlaGlu 185  
 Db 2 GAACAGCCCAAGCTCCATGGGATGATTCGGACTGAAATAGCGAGCGAAATTCGCCAA 61  
 QY 186 GluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGluTyLysLeuLysVal 205  
 Db 62 GAGATGGAAAAAGAGAGCGCTTCTTCAGCTACAGATGTGCGAGTATCTCTGGAAGTTC 121  
 QY 206 AsnGluLysIleLysLysGlyValAspLeuLeuGlnAsnLeuLysLysTyrrPheHis 225  
 Db 122 AACGAAATCAAGATTAAGAGGGAGTAGATTTACTTCGAATCTGATCAATATCTTCTAT 181  
 QY 226 AlaGlnCysAsnPhenGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIle 245  
 Db 182 GCCCAATGCAATTTTTTTCAGATGGACTCAAGCGGTGGAAGCCTCAACCTTCCATT 241  
 QY 246 GluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGln 265  
 Db 242 GAAACGCTGTCTACGGATCTTACACAGCATCAACAGGCCAGGATGAAGAAGAGGCGAG 301  
 QY 266 LeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSer 285  
 Db 302 TTGATACAGCTTCAGATATTTTGAATCCCATTCAGGTTGAACAGAAAGAGGACTCC 361  
 QY 286 GlnIleArgGlnSerThrAlaTySerLeuHisGlnProGlnGlyAsnLysGluHisGly 305  
 Db 362 CAAATTCGTACAGACACAGCTTATAGCTTACATCAGCTCAGGAAACAAGAACATGGG 421  
 QY 306 ThrGluArgAsnGlySerLeuTyLysLysSerAspGlyIleArgLysValTrpGlnLys 325  
 Db 422 ACCGAGCGAAGCGAGCGCTCTACAGAAGAGTACGCGGATCCGAAAAAGTGTGGCAGAA 481  
 QY 326 ArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAla 342  
 Db 482 AGGAATGTTTCAGTTAAATGTTTCTGACCATATCCCATGTGTACCGCT 532  
 RESULT 4  
 LOCUS BUI55022 846 bp mRNA linear EST 03-SEP-2002  
 DEFINITION AGENCOURT\_7944696 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6148570  
 5', mRNA sequence.  
 ACCESSION BUI55022  
 VERSION BUI55022.1 GI:22668554  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 846)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DTF  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM13480 Row: d Column: 11  
 High quality sequence stop: 524.  
 Location/Qualifiers  
 1..846  
 /organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6148570"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Alignment Scores:
Pred. No.: 4.57e-162 Length: 846
Score: 165.00 Matches: 165
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.40% Indels: 0
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x BU155022 (1-846)
Qy 535 TyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVal 554
Db 26 TACGCAAGGAAGACGCGGATTAACCGCGGAAGCTTTCAGATCTTTGGAGCGGTC 85
Qy 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574
Db 86 AAAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGGTGTGGATCTTACGGAA 145
Qy 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594
Db 146 AAAATCCCACTGGCCACGACATGAGCGCGATGAACGGCCCTCCACCTTGCAGTCAGA 205
Qy 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614
Db 206 TCCGTGGATCGAACCCTCTCTTCACATTGTAGACTTTTATGTTTCAGAACAGTGGGAACCTG 265
Qy 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAla 634
Db 266 GATAAACAGACAGGAAAGGACGACAGCCCTGCCTACTACTGCTGCTGACCGCAATGCC 325
Qy 635 GluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
Db 326 GAGTGCCCTCAAGTGTCTCTCGGGGAAGGCTCCATCGAGATAGCAACAGATCAGGA 385
Qy 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuThr 674
Db 386 GAGACTCCGCTGGACATTGCCAAGCGCTCAAGCAGAGCACTGTGAGGAGCTGCTGACC 445
Qy 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluThrArgLeuLeu 694
Db 446 CAAGCCTTATCTGGAAGATTTAATTTCTCACGTTTCCAGTTTGAATGAATGGCGACTACTC 505
Qy 695 HisGluAspLeuAsp 699
Db 506 CACGAAGACCTGGAT 520

RESULT 5
BU703510
LOCUS
DEFINITION UI-M-F00-bzo-m-08-0-UI.r1 NIH_BMAP_F00 Mus musculus CDNA clone
IMAGE:6405295 5', mRNA sequence.
ACCESSION BU703510
VERSION BU703510.1 GI:23629426
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.
Location/Qualifiers
1..689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6405295"
/tissue_type="whole brain"
/dev_host="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F00"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TGAGAGAGCC. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN
Alignment Scores:
Pred. No.: 6.83e-157 Length: 689
Score: 160.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 15.90% Indels: 2
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x BU703510 (1-689)
Qy 312 LeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331
Db 99 CTCTACAAGAAGACGATGGGATCCGGAAGTGTGGCAGAGAGAGAAAGTGTTCGCTTAA 158
Qy 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
Db 159 AATGGCTTCTTCACCATCTCCACGGCACTGCCAACCGGCCCTGCGAAGCTTCACTG 218
Qy 352 LeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHis 371
Db 219 CTAACTGCCAGGTGAAGACCAATCTCGAGAGAAGTGTTCGACCTCATATCAT 278
Qy 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTyrMetSerVal 391
Db 279 GACAGGACGTACCACTTCCAAAGCGAAGACCAAGAAATGTCAGATATGATGTCGTA 338
Qy 392 LeuGlnAsnSerLysGluAlaLeuAsnAlaPheLysGlyAspAsnThrGly 411
Db 339 CTGCAAGACCAAGGAAGAGAGCTCTGAACACGCCCTTAAAGGGTGTATGACACACTGGA 398
Qy 412 GluAsnAsnIleValGlnLeuThrLysGluIleSerGluValGlnArgMetThr 431

```



```

Db      399 GAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGAGGTGCGAGGATGACG 458
Qy      432 GlyAenAspValCysCysGlyAlaProAspProThrTrpLeuSerThrAenLeu 451
Db      459 GGCATGACGTGTGCTGCGACTGTGGGACCAAGGATCCGAGCTGCTCTTACCAACCTG 518
Qy      452 GlylleLeuThrCysIleGluCysSerGlylleHisArgGluLeuGlyValHisTrpSer 471
Db      519 GGCATCCCTGACTTGCATCGAGTGTCTGGGATCCACCGGAGCTGGGGGTTCATTACTCC 578
Qy      472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAs 491
Db      579 AG-GATGAGTGCCTGACATTAGATGATGGGAACGCTGAGCTTCCTGCTGCCAGAA 637
Qy      491 nileGlyAenAlaGlyPheAenGluLeuMetGluCysCysLeuPro 506
Db      638 TATCGGAATGCAGGCTTTAATGAAATTATGGAGTGTTCCTACCG 683

RESULT 6
CN538273      731 bp      mRNA      linear      EST 29-APR-2004
LOCUS      UI-M-H50-qgr-k-12-0-UI.r1 NIH_BMAP_H50 Mus musculus cDNA clone
DEFINITION      IMAGE:30675611 5', mRNA sequence.
ACCESSION      CN538273
VERSION      CN538273.1 GI:46866429
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 731)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 731
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30675611"
/tissue_type="Upper Head"
/dev_stages="embryo 9.5 - 10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_H50"
/note="Organ: Upper Head; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CCAACTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)".

FEATURES
source

ORIGIN
Alignment Scores:

```

```

Pred. No.:      7,23e-157      Length:      731
Score:          160.00      Matches:      192
Percent Similarity: 98.97%      Conservative: 0
Best Local Similarity: 98.97%      Mismatches: 1
Query Match:    15.90%      Indels:      2
DB:              7      Gaps:      0

US-09-914-042-1 (1-1006) x CN538273 (1-731)

Qy      312 LeuTrpLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331
Db      152 CTCTACAAGAGAGCGATGGATCCGGAAGTGTGGCAGAGAGGAAGTGTTCGCTTAA 211
Qy      332 AenGlyPheLeuThrIleSerHisGlyThrAlaAenArgProProAlaLysLeuAenLeu 351
Db      212 AATGGCTTCCTCACCATCTCCACGGCACTGCCAACCGCCACCTGCCAAGCTCAACCTG 271
Qy      352 LeuThrCysGlnValLysThrAenProGluLysLysCysPheAspLeuIleSerHis 371
Db      272 CTAACTCTGCCAGGTGAAGACCAATCTCTGAGAGAGAAAGTGTTCGACCTCATATCACAT 331
Qy      372 AspArgThrTrpHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerVal 391
Db      332 GACAGGACGTACCACTTCCAGCGGAGACGACGACAGATGTTCAGATATGATGTCTGTA 391
Qy      392 LeuGlnAenSerLysGluAlaLeuAenAenAlaPheLysGlyAspAenThrGly 411
Db      392 CTGCAGAACCAAGGAAGAGCTCTGAACAACGCTTTAAGGGTGTATGACCAACACTGGA 451
Qy      412 GluAenAenIleValGlnGluLeuThrLysGluIleSerGluValGlnArgMetThr 431
Db      452 GAAATAACATCGTCCAGAGCTGACCAAGAGAGATCATCTCGAGGTGCGAGGATGACG 511
Qy      432 GlyAenAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAenLeu 451
Db      512 GGCATGACGTGTCTGGCACTGTGGGCACACAGATCCGAGCTGGCTCTTACCAACCTG 571
Qy      452 GlylleLeuThrCysIleGluCysSerGlylleHisArgGluLeuGlyValHisTrpSer 471
Db      572 GGCATCTCTGACTTGCATCGAGTGTCTGGGATTACCGGGAGCTGGGGGTTCATTACTCC 631
Qy      472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491
Db      632 AG-GATGAGTCCCTGACATTAGATGTATTGGGAACGCTCGAGCTTCTGCTTCCCAAGAA 690
Qy      491 nileGlyAenAlaGlyPheAenGluIleMetGluCysCys 504
Db      691 TATCGGAATGCAGGCTTTAATGAAATTATGGAGTGTTCG 730

RESULT 7
CN460643      773 bp      mRNA      linear      EST 21-APR-2004
LOCUS      UI-M-HB0-coy-a-05-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
DEFINITION      IMAGE:30652708 5', mRNA sequence.
ACCESSION      CN460643
VERSION      CN460643.1 GI:46466369
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 773)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html

```

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

# FEATURES

Location/Qualifiers  
1. .773  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30652708"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_HB0"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATTGAAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Alignment Scores:  
Pred. No.: 7,638-157 Length: 773  
Score: 160.00 Matches: 194  
Percent Similarity: 98.98% Conservative: 0  
Best Local Similarity: 98.98% Mismatches: 1  
Query Match: 15.90% Indels: 2  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CM460643 (1-773)

Qy 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerVallys 331  
Db 53 CTCTACAGAAGAGCGGATGGGATCGGAAGTGTGGCAGAGAGAGAGTTCGGTTAAA 112  
Qy 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351  
Db 113 AATGGCTTCTCACCATCTCCACGGCACTGCCAACCGGCCACCTGCCAAGCTCAACCTG 172  
Qy 352 LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371  
Db 173 CTACCTGCCAGGTGAGACCATCTCTGAGGAGAGAGAGTGTTCGACCTCATATCAT 232  
Qy 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391  
Db 233 GACAGGAGTACCCTTCCAGCGGAAGACGAACAAAGATGTCAGATGATGATGCTGTA 292  
Qy 392 LeuGlnAsnSerLysGluLysAlaLeuAsnAlaPheLysGlyAspAsnThrGly 411  
Db 293 CTGCAGAACAGCAAGAGAAAGCTCTGAACACGCTTTAAAGGTGATGACAACTGGA 352  
Qy 412 GluAsnAsnIleValGlnLeuThrLysGluIleIleSerGluValGlnArgMetThr 431  
Db 353 GAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGGATGACG 412  
Qy 432 GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451  
Db 413 GGCAATGACGTGTCTGCGACTGTGGGCACCCAGATCCGAGTGGCTCTCTACCAACCTG 472  
Qy 452 GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSer 471  
Db 473 GGCATCTCTGACTGTCAGTGTCTGGGATCCACCGGGAGCTGGGGGTTCATTACTCC 532  
Qy 472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491

Db 533 AG-CATCAGTCCCTGACATTAGATGATTATGGGAACGCTCTGAGCTTCTGCTTCCAGAA 591  
Qy 491 nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuPro 506  
Db 592 TATCGGAATGCAGGCTTTAATGAAATATATGAGTGTTCCTACCG 637

## RESULT 8

CO429713 781 bp mRNA linear EST 06-JUL-2004  
LOCUS UI-M-HX0-crw-p-02-0-UI.r1 NIH\_BMAP\_HX0 Mus musculus cDNA clone  
DEFINITION IMAGE:30686473 5', mRNA sequence.  
ACCESSION CO429713  
VERSION CO429713.1 GI:49676007  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 781)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaaba-remail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1. .781  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30686473"  
/tissue\_type="whole eye"  
/dev\_stage="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_HX0"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATAATATACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Alignment Scores:  
Pred. No.: 7,71e-157 Length: 781  
Score: 160.00 Matches: 194  
Percent Similarity: 98.98% Conservative: 0  
Best Local Similarity: 98.98% Mismatches: 1  
Query Match: 15.90% Indels: 2  
DB: 7 Gaps: 0  
US-09-914-042-1 (1-1006) x CO429713 (1-781)  
Qy 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerVallys 331

```

Db      83  CTCTACAGAGAGCGATCGGATCCGAAAGTGTGCAGAGAGGAAGTTCCTGTTAAA 142
Qy      332  AaNGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
Db      143  AATGCTTCTCCATCCATCTCCACGGCACTGCCAACCGGCCACCTGCCCAAGCTCAACCTG 202
Qy      352  LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371
Db      203  CTAACCTGCCAGGTGAAGACCAATCTCGAGGAGAAAGATGTTCGACCTCATATCACAT 262
Qy      372  AspArgThrTyHisPheGlnAlaGluAspGluGlnCysGlnIleTyrMetSerVal 391
Db      263  GACAGACGTACCATCTTCAGCGGAAGACGAACGAATGTGCATGTGATGTCTGTA 322
Qy      392  LeuGlnAsnSerLysGluGluAlaLeuAsnAlaPheLysGlyAspAsnThrGly 411
Db      323  CTGCAGAACAGCAAGGAAGAGCTCTGAACAACGCCCTTTAAGGTGTGATGACACACTGA 382
Qy      412  GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
Db      383  GAAATAACATCGTCAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGAGGATGACG 442
Qy      432  GlyAsnAspValCysCysAspCysGlyAlaProAspProThrLeuSerThrAsnLeu 451
Db      443  GCAATGACGTGTGTGCGACTGTGGGCGACGACGATCCGAGTCTCTTACCAACCTG 502
Qy      452  GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSer 471
Db      503  GGCATCTGACTTGGATCGATCGATGCTCTGGATCCACCGGGAGCTGGGGTTCATTACTCC 562
Qy      472  Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491
Db      563  AG-GATGCAGTCCCTCGACATTAGATGTATTGGAACTGTGAGTCTCTGCTTGTGCCAAGA 621
Qy      491  nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuPro 506
Db      622  TATCGGGAATGACGCTTTAATGAATAATTATGGAGTGTTCCTTACCG 667

```

```

RESULT 9
BP280849 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
LOCUS
DEFINITION
BP280849
VERSION
BP280849.1 GI:52194581
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

```

```

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

AUTHORS
Suzuki, Y., Yamaehita, R., Shiota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.

```

```

TITLE
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions

```

```

JOURNAL
Genome Res. 14 (9), 1711-1718 (2004)

```

```

COMMENT
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.

```

```

FEATURES
source
1..582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KGR09051"
/cell_line="KG-1-C"
/clone_lib="Sugano cDNA library, KG-1-C"
/note="Glioma"

```

```

ORIGIN

```

```

Alignment Scores:
Pred. No.: 8.32e-154 Length: 582
Score: 157.00 Matches: 157
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatch: 0
Query Match: 15.61% Indels: 0
DB: 5 Gaps: 0

```

```

US-09-914-042-1 (1-1006) x BP280849 (1-582)

```

```

Qy      43  AlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAlaIleAen 62
Db      112  GCTTGGACGTGGACCGGATGTTCTTTACAAATGAAGAAATCCGTGAAGCAATCAAC 171
Qy      63  SerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGluLysPhe 82
Db      172  AGCTCTGGCTGGCTCACGTGAAATGAAGAGCAGTACACCCAGGCTCTGGAGAAGTTT 231
Qy      83  GlyLysAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLysPheSer 102
Db      232  GCGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAGTGGCTTCTCGAAGTTCTCA 291
Qy      103  ValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAenIle 122
Db      292  GTGTTTCAAAAGGAGTTGACAGCACTTTTCAAAACCTGTATCAGATATGAACACATA 351
Qy      123  IleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGlyAspLeu 142
Db      352  ATCTCTTCCCTTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTGAAGGGGATCTG 411
Qy      143  LysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIleGluLys 162
Db      412  AAAAAGCCTTTGTATAAGCTTGGAAAGACTATGAACAAATAAACCAAGATAGAAAG 471
Qy      163  GluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGlu 182
Db      472  GAGAAAAGGAAACACGCCAAGCTCATGGGATGATTCGAGCTGAAATGAAGCGAGCGGAA 531
Qy      183  IleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCys 199
Db      532  ATTCCGAGAGATGGAAAGAGAGGCGCTTCTTCCAGCTACAGATGTGC 582

```

```

RESULT 10
BM839307
LOCUS

```

```

DEFINITION
K-EST0116049 S9SNU601 Homo sapiens cDNA clone S9SNU601-68-A09 5',

```

```

ACCESSION
BM839307
VERSION
BM839307.1 GI:19195716
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

```

```

ORGANISM
Homo sapiens

```

```

REFERENCE
1 (bases 1 to 485)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

AUTHORS
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.

```

```

TITLE
21C Frontier Korean EST Project 2001

```

```

JOURNAL
Unpublished (2002)

```

```

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

```

```

Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr

```

```

Plate: 68 row: A column: 09
High quality sequence stop: 485.

```

```

FEATURES
Location/Qualifiers
1..485
/organism="Homo sapiens"
/mol_type="mRNA"

```

```

/db_xref="taxon:9606"
/clone="S9SNU601-68-A09"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10P"
/clone_lib="S9SNU601"
/note="Organ: Stomach; Vector: pME18-FL3; Site: 1: XhoI;
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII- digested pME18S-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 7,87e-153 Length: 485
Score: 156.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.51% Indels: 0
DB: 4 Gaps: 0

```

US-09-914-042-1 (1-1006) x BM839307 (1-485)

```

QY 557 ArgAspIlePheGlyLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysile 576
DB 3 AGAGATATTTTGGATTGCTCCAGCTATGCTGATGTTGGATCTTACGGAAAAATC 62
QY 577 ProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVal 596
DB 63 CCATGGGCCAACGGGACATGAGCGGATGAACGGCCCTCCACCTTGCAGTCAGATCCGTG 122
QY 597 AspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLys 616
DB 123 GATCGAACCCTCTTTCACATTGTAGACTTTTGTAGTTCAGAACAGTGGGAACCTGGATAAA 182
QY 617 GlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCys 636
DB 183 CAGACAGGGAAAGGAGCAGACGCCCTGCTACTGCTGCTGACCGACAATGCCGAGTGC 242
QY 637 LeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThr 656
DB 243 CTCAGATTGCTCTCTCGGGGGAAGCCCTCCATCGAGATAGCAACGAGTCCAGGAGACT 302
QY 657 ProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuThrGlnAla 676
DB 303 CGCTGGGACATTGCCAAGCGCCTCAAGCACGAGCAGCTGTGAGGAGCTGCTGACCCAGCC 362
QY 677 LeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGlu 696
DB 363 TTATCTGGAGAGATTAAATCTCAGCTTCACGTTTCAGTATGATGATGCGGACTACTCCAGNA 422
QY 697 AspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSer 712
DB 423 GACCTGGATGAAAGTGATGACGACATGGATGGATGAGAAATTGCAGCCCACT 470

```

## RESULT 11

```

CF532249
LOCUS
DEFINITION
IMAGE:30356727 5', mRNA sequence.
ACCESSION

```

```

CF532249.1 GI:34584217
EST.
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
REFERENCE
1 (bases 1 to 702)
AUTHORS
TITLE
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..702
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30356727"
/tissue_type="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GHO"
/note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
Site 2: Not I; the library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAAGTCAAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

```

## FEATURES

source

## ORIGIN

```

Alignment Scores:
Pred. No.: 1.43e-150 Length: 702
Score: 154.00 Matches: 188
Percent Similarity: 98.95% Conservative: 0
Best Local Similarity: 98.95% Mismatches: 1
Query Match: 15.31% Indels: 2
DB: 7 Gaps: 0
US-09-914-042-1 (1-1006) x CF532249 (1-702)
QY 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrile 337
DB 3 GGGATCCGGAAGAGTGTGCAGAGGAGGAGTGTTCGGTTAAATAATGGCTTCTCCACCATC 62
QY 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
DB 63 TCCACGCGCACTGCCACCGCCACCTGCCAAGCTCAACCTGCTAACCTGCAGGTGAAG 122
QY 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
DB 123 ACCAATCTCTGAGGAGAGAGAGTGTTCGACCTCATATCATGACAGGACGTACCATCTTC 182
QY 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397

```

```

Db      183  CAGCGGAGACCAACAAGATGTCAGATATGGATGTCTGTACTCCAGAACAGCAAGGAA 242
Qy      398  GluAlaLeuAsnAlaPheLysGlyAspAsnThrGlyGluAsnAlaValGln 417
Db      243  GAAGCTCTGAACAACGCTTTAAAGGATGATGACACACTGGAGAAAATAACATCTGCCAA 302
Qy      418  GluLeuThrLysGluLeuSerGluValGlnArgMetThrGlyAsnAspValCys 437
Db      303  GAGCTGACCAAGGAGATCATCTCGAGAGGTGCAGAGGATGACGGCAATGACGTGTGCTGC 362
Qy      438  AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyLeuThrCys 457
Db      363  GACTGTGGGACCAAGATCCGACGTGGCTCTCTACCAACCTGGGCATCTGACTTGGCATC 422
Qy      458  GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeu 477
Db      423  GAGTCTCTGGGATCCACCGGAGCTGGGGTTTCAATTACTCCAG-GATGCAGTCCCTGAC 481
Qy      477  rleuaspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGly 497
Db      482  ATTAGATGATTGGGAAGCTCTGAGCTTCTGCTTCCCAAGAATATCGGGAATGCAAGGCTT 541
Qy      497  eAsnGluLeuMetGluCysCysLeuPro 506
Db      542  TAATGAATAATGGAGTGTGCTACCG 569

RESULT 12
LOCUS   BQ439046 884 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT_7761529 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020058
5', mRNA sequence.
ACCESSION BQ439046
VERSION   BQ439046.1 GI:21178122
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 884)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgabbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM13223 row: i column: 19
           High quality sequence stop: 733.
           Location/Qualifiers
             1..884
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:6020058"
               /tissue_type="epithelioid carcinoma"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH MGC 70"
               /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
               Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
               Average insert size 1.1 kb. Library constructed by Life
               Technologies."

FEATURES
source
1..884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6020058"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

ORIGIN
Alignment Scores:
Pred. No.: 1,78e-150 Length: 884
Score: 154.00 Matches: 242
Percent Similarity: 99.18% Conservative: 0
Best Local Similarity: 99.18% Mismatches: 1

```

```

Query Match: 15.31% Indels: 2
DB:          5          Gaps: 0
US-09-914-042-1 (1-1006) x BQ439046 (1-884)

Qy      318  GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db      13  GGGATCCGAAAGATGTGGCAGAAAAGGAATGTTTCAGTTAAANAATGGTTTCTTGACCAT 72
Qy      338  SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db      73  TCCCATGGTACCGCTAACCGGCTCTCTGCAAGGCTCAACCTGCTAACCTGCCAGGTGAAG 132
Qy      358  ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
Db      133  ACCAACCTGAGGAGAGAGTGTGTTGACCTTATTTTACATGACAGAACTTACCACCTTT 192
Qy      378  GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGluAsnSerLysGlu 397
Db      193  CAAGCTGAAGATGAACAGGAATGTCAAATATGGATGTCTGTCTGCTCAAAATAGCAAGAA 252
Qy      398  GluAlaLeuAsnAlaPheLysGlyAspAsnThrGlyGluAsnAlaValGln 417
Db      253  GAAGCTTTAAACAATGCATTTTAAAGGGATGACAACTACTGGAGAAAATAACATCGTCCA 312
Qy      418  GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCys 437
Db      313  GAACCTGACAAAGGAGATCATCTCAGAAGTCAGAGGATGACGGCAATGACGCTGCTGT 372
Qy      438  AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCys 457
Db      373  GACTGTGGGGCCGACAGATCTTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATC 432
Qy      458  GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeu 477
Db      433  GAGTGTTCGGGAATCCACCGAGAGCTGGGGTTTCAATTCTCAG-GATGCAGTCCCTGAC 491
Qy      477  rLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGly 497
Db      492  CTTAGATGATCTGGGAACATCTGAGCTGCTCGCCAAAGAATATTGGGAATGCAAGCTT 551
Qy      497  eAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGly 517
Db      552  TAATGAGATCATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCAACCCAGGAG 611
Qy      517  rAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTgTyrAla 537
Db      612  CGCATGAATCAAGAAGGACTACATCACAGCCCAAGTACATCGAGAGAGATACGCAAG 671
Qy      537  GlyLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuGluAlaValLysThr 557
Db      672  GAAGAGCACCGGGATTAACGCGCGAGCTTACAGATCTTTTGGGAGCGCGTCAAAACGAG 731
Qy      557  gAspIlePhe 560
Db      732  AGATATTTTT 741

RESULT 13
LOCUS   CD358314 948 bp mRNA linear EST 29-MAY-2003
DEFINITION AGENCOURT_14275993 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30388565 5', mRNA sequence.
ACCESSION CD358314
VERSION   CD358314.1 GI:31129749
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 948)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)

```

## COMMENT

Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgapbs-r@mail.nih.gov  
cDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM460 row: k column: 06  
High quality sequence stop: 543.  
Location/Qualifiers  
1. .948  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30388565"  
/lab\_host="DH10B-Ton A ( T1 and T5 phage resistances)"  
/clone\_lib="NIH\_MGC\_180"  
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;  
Site 2: EcoRV (destroyed); Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.68 kb. Library was  
constructed by (Invitrogen). Note: this is a NIH\_MGC  
Library."

## FEATURES

source

## ORIGIN

Alignment Scores:  
Pred. No.: 2,738-147 Length: 948  
Score: 151.00 Matches: 179  
Percent Similarity: 98.90% Conservative: 0  
Best Local Similarity: 98.90% Mismatches: 1  
Query Match: 15.01% Indels: 2  
DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x CD358314 (1-948)

QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
DB 3 AAAGTGTGCAGAAAAGGAATGTTTCAGTTAAATGTTTCTGCACCATATCCCATGTT 62  
QY 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
DB 63 ACCGCTAACCGGCCCTCTGCAAAAGTCAACCTGTCAACCTGCAGGTGAAGACCAACCT 122  
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
DB 123 GAGGAGAGAGAGTGTTCAGCTTATTTTCACATGACAGACTTACCATCTTCAAGCTGAA 182  
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
DB 183 GATGAACAGGAATGTCMAATATGGATGTCGTGCTGCAAAATAGCAAGAAAGACTTTA 242  
QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
DB 243 AACAAATGCAATTTAAAGGGGATGACAATACTGGAGAAAATAACATCGTCCAAAGAACTGACA 302  
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
DB 303 AAGGAGATCATCTCAGAGTGCAGAGGATGACGGGCAATGACGTCTGTGCTGACTGTGGG 362  
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
DB 363 GCGCAGATCTCTACATGGCTTTCCACCAACCTGGGCATCTCAGCTGCATCAGGTGTTCC 422  
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVa 480  
DB 423 GGAAATCCACCGAGAGCTGGGGGTTTCATTATTCAG-GATGCAAGTCCCTGACCTTAGATGT 481  
QY 480 IleuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500

Db 482 ACTGGGAACATCTGAGCTGCTGCTCGCAAGAAATATGGGAATCAGGCTTTAATGAGAT 541  
QY 500 e 500  
DB 542 C 542  
RESULT 14  
CF745192  
LOCUS  
DEFINITION  
CF745192 714 bp mRNA linear EST 10-OCT-2003  
IMAGE:30620594 5', mRNA sequence.  
ACCESSION  
CF745192  
VERSION  
CF745192.1 GI:37641532  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 714)  
AUTHORS  
NIH-MGC http://mgs.nci.nih.gov/.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
seq primer: pYX-5.  
Location/Qualifiers  
1. .714  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30620594"  
/tissue\_type="whole brain"  
/dev\_stage="1.5, and 15 days newborn"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_GVO"  
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated with  
EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CGAAGTGAAT. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

ORIGIN  
Alignment Scores:  
Pred. No.: 2,986-144 Length: 714  
Score: 148.00 Matches: 148  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 14.71% Indels: 0  
DB: 7 Gaps: 0  
US-09-914-042-1 (1-1006) x CF745192 (1-714)  
QY 312 LeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331

```

Db      90  CTCTACAGAGAGGATGGATCCGAAAGTGTGGCAGAGAGAGAGTTCCTTTAAA 149
Qy      332  AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeu 351
Db      150  AATGCTTCTTCAACCATCTCCACAGGCACTGCCAACCGGCCACCTGCCCAAGCTCAACCTG 209
Qy      352  LeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHis 371
Db      210  CTAACCTCCAGGTGAAGACCAATCTCGAGGAGAGAAAGTGTTCGACCTCATATCACAT 269
Qy      372  AspArgThrTyrHisPheGluAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391
Db      270  GACAGAGCTACCATCTCCAGCGGAGACCAAGAAATGTCAGATGATGATGTCGTGA 329
Qy      392  LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411
Db      330  CTGCAGAACACAGCAAGAGAGCTCTGAACAACGCCCTTTAAGGTGATGACACACTGGA 389
Qy      412  GluAsnAsnIleValGlnGluLeuThrLysGluLleIleSerGluValGlnArgMetThr 431
Db      390  GAAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGAGTACG 449
Qy      432  GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451
Db      450  GGCAATGACGTGTCTGCGACTGTGGGCAACAGATCCGAGCTGCGTCTCTTACCAACCTG 509
Qy      452  GlyIleLeuThrCysIleGluCys 459
Db      510  GGCATCTGACTTGCATCGATGTC 533

```

## RESULT 15

CN535077

LOCUS

DEFINITION

UI-M-HS0-cqi-a-11-0-UI-r1 NIH\_BMAP\_H50 Mus musculus cDNA clone

IMAGE:30674218 5', mRNA sequence.

ACCESSION

CN535077

VERSION

EST.

CN535077.1 GI:46863233

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 742)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

<http://genome.uiowa.edu/distribution/mousefi.html>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5

Location/Qualifiers

1..742

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:30674218"

/tissue\_type="Upper Head"

/dev\_stage="embryo 9.5 - 10.5 dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_H50"

/note="Organ: Upper Head; Vector: pYX-Asc; Site\_1: EcoR I; Site\_2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Alignment Scores: 3.09e-144 Length: 742  
 Pred. No.: 148.00 Matches: 182  
 Score: 148.00  
 Percent Similarity: 98.91% Conservative: 0  
 Best Local Similarity: 98.91% Mismatches: 1  
 Query Match: 14.71% Indels: 2  
 DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN535077 (1-742)

```

Qy      324  GlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsn 343
Db      69  CAGAAGAGGAAGTGTTCGTTTAAATGGCTTCTCACCATCTCCACGGCACTGCCAAC 128
Qy      344  ArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGluLys 363
Db      129  CGGCCACCTGCCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAATCTCTGAGGAGA 188
Qy      364  LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
Db      189  AAGTGTTCGACCTCATATCATGACAGGACGTACCACTTCCAGCGGAAGACGACAA 248
Qy      384  GluCysGlnIleTrpMetSerValLysGlnAsnSerLysGluGluAlaLeuAsnAla 403
Db      249  GAATGTCTAGATATGATGTCTGTACTGCAGAACAGCAAGGAGAGAGCTCTGAAACAC 308
Qy      404  PheLysGlyAspAspAsnThrGlyLysAsnIleValGlnGluLeuThrLysGluIle 423
Db      309  TTTAAGGTGTATGACACACACTGGAGAAATTAACATCGTCCAAGAGCTGACCAAGGAGAT 368
Qy      424  IleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAsp 443
Db      369  ATCTCGAGGTGCAGAGGATGACGGCAATGACGTGTCTGCGACTGTGGGGCACCAGAT 428
Qy      444  ProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHis 463
Db      429  CCGACGTGGCTCTCTTACCAACCTGGGCATCTGACTTGCATCGAGTGTCTCTGGGATCCAC 488
Qy      464  ArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValLeuGlyTh 483
Db      489  CGGAGCTGGGGGTTCATTACTCCAG-GATCGAGTCTCTGACATTAGATGATTGGGAAC 547
Qy      483  rSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCys 503
Db      548  GTCTGAGCTTCTGCTTGCAGAAATATCGGAATATCGGAGTTCAGGCTTTAATGAATAT 607
Qy      503  sCysLeuPro 506
Db      608  TTGCTTACCG 617

```

Search completed: August 5, 2005, 00:18:10  
 Job time : 7514 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 22:16:06 ; Search time 1433 Seconds  
(without alignments)  
4550.746 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDGPKRGKAPVPSFVHFAD 1006

Scoring table:

Oligo Xgapop 60.0 , Xgapext 60.0  
Xgapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7297361 seqs, 3241162794 residues

Word size: 1

Total number of hits satisfying chosen parameters: 14583226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool\_h/US0914042/runat\_04082005\_090749\_2764/app\_query.fasta\_1.1159  
-DB=Published\_Applications\_NA -QPM=fastap -SUFFIX=oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US0914042 @CCN 1 1 658 @runat\_04082005\_090749\_2764  
-NCPUI=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/1/pubna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/1/pubna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*  
23: /cgn2\_6/ptodata/1/pubna/US11A\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/1/pubna/US11\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*  
26: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	717	71.3	5711	21	US-10-956-157-1144
2	110	10.9	3123	17	US-10-104-047-314
3	110	10.9	3436	11	US-09-764-875-37
4	79	7.9	3413	11	US-09-764-875-358
5	78	7.8	442	10	US-09-318-995-27010
6	48	4.8	563	9	US-09-815-343-1154
7	48	4.8	563	18	US-10-097-105-1154
8	38	3.8	958	13	US-10-027-632-34151
9	38	3.8	958	17	US-10-027-632-34151
10	30	3.0	409	18	US-10-276-774-404
11	30	3.0	3974	18	US-10-210-281-73
12	22	2.2	5582	17	US-10-062-674-1606
13	19	1.9	60	10	US-09-908-975-22014
14	17	1.7	326	9	US-09-815-343-1151
15	17	1.7	326	18	US-10-097-105-1151
16	16	1.6	469	18	US-10-276-774-288
17	16	1.6	1523	21	US-10-764-420-1028
18	14	1.4	1038	18	US-10-276-774-639
19	13	1.3	167	9	US-09-864-761-30407
20	13	1.3	323	9	US-09-728-446-1051
21	13	1.3	446	9	US-09-864-761-13841
22	12	1.2	296	20	US-10-333-872A-161
23	12	1.2	399	17	US-10-442-535A-21815
24	12	1.2	399	18	US-10-085-783A-21815
25	12	1.2	4211	17	US-10-172-118-1101
26	12	1.2	4211	18	US-10-342-887-1101
27	12	1.2	28055	17	US-10-085-117-94
28	11	1.1	451	17	US-10-062-674-482
29	10	1.0	417	9	US-09-912-020-140
30	10	1.0	417	20	US-10-771-241-140
31	10	1.0	697	9	US-09-912-020-446
32	10	1.0	697	20	US-10-027-632-157073
33	10	1.0	824	13	US-10-027-632-110267
34	10	1.0	824	17	US-10-027-632-110268
35	10	1.0	2339	13	US-10-027-632-110268
36	10	1.0	2339	13	US-10-027-632-110268
37	10	1.0	2339	13	US-10-027-632-110267
38	10	1.0	2339	17	US-10-027-632-110267
39	10	1.0	2339	17	US-10-027-632-110268
40	10	1.0	2339	17	US-10-027-632-110268
41	10	1.0	3812	17	US-10-104-047-981
42	10	1.0	4050	21	US-10-490-605-1
43	9	0.9	201	19	US-10-741-601-21882
44	9	0.9	307	21	US-10-487-561-51
45	9	0.9	317	19	US-10-437-963-21619

# ALIGNMENTS

## RESULT 1

US-10-956-157-1144  
; Sequence 1144, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956.157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ IDS NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1144  
; LENGTH: 5711

```
! TYPE: DNA
! ORGANISM: Homo sapiens
US-10-956-157-1144

Alignment Scores:
Pred. No.: 0 Length: 5711
Score: 717.00 Matches: 1003
Percent Similarity: 99.41% Conservative: 0
Best Local Similarity: 99.41% Mismatches: 3
Query Match: 71.27% Indels: 6
DB: 21 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAla 20
Db 341 ATGCGGACCAAGATCTCGTGTGGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG 400

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCAGCGCTCCAGCTTCACACCCGACGCGGCGAGTCCGCGNACACTGTGGCGGCATC 460

Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrIysMetLysLysSerValIysAla 60
Db 461 GAGGAGGCTTTGGACGTGGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACACCCAGGCTCTGGAG 580

Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCACTGTGTATGACAGATGACCCAGATTTAGGAAGTGGCTTCTCGAAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAAATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
Db 701 AACATAATCTCTTCCCTTTGGACAGTTTGTGAAGGGGGACCTGAAAGGAGTGAAGGG 760

Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrIysIle 160
Db 761 GATCTGAAAAGAGCTTTTGATTAAGCTTGGAAAGGACTATGAAACAAAATAAACCAAGATA 820

Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAGGAGAAAAGGAAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATAGCGGA 880

Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGAAAAGGAGAGGGCGCTTCTTCCAGCTACAGATGTGCGAG 940

Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGCTGAGGTCAAGAAATCAAGATTAAGAAAGGAGTAGATTTACTTCAGAAATCTG 1000

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTTCATGCCCAATGCAATTTTTTTTTCAGGATGGACTCAAAAGCGGTGAAAGC 1060

Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACCTTCATTGAAACGCTGTCTACGGATCTTCAACGATCAAAACAGGCCCCAGGAT 1120

Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGACGTTGATCAGCTTCGAGATATTTTGAATTCGCGATTTCAGAGTTGAA 1180

Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240
```

```
Qy 301 AsnLysGluHisGlyThrGluAtqAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAAAGGAACATGGGACCGGACCGGACCGCTCTACAAAGAGAGTGACGGGATCCGA 1300

Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAAGGAAATGTTCAAGTTAAAAAATGGTTTCTGACCATATCCATGGT 1360

Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGAACCTGCCAGGTGAAGACCAACCT 1420

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAGAAGTCTTTTGACCTTATTTACATGACAGAACTTACCACTTTCAAGCTGAA 1480

Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATGAACAGGAATGTCAAAATATGATGCTGTGCTGCAAAATAGCAAAAGAAAGCTTTA 1540

Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACAAATGCAATTTAAGGGGATGACAAATCTGGAGAAAATAAACATCGTCCAAGAACTGACA 1600

Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGG 1660

Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTCTGACCTCGAGTGTTC 1720

Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVa 480
Db 1721 GGAATCCACCAGAGAGCTGGGGGTTTCAATATTTCCAG-GATGCAGTCCCTGACCTTAGATGT 1779

Qy 480 lLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1780 ACTGGGAACATCTGAGCTGTGCTGCCAAGAAATATTTGGGNATGCGAGGCTTATAGAT 1839

Qy 500 eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs 520
Db 1840 CATGGAATGTTGCTCTACCGCTGAGGACTCAGTCAAAACCCCAACCCAGCAGCGACATGAA 1899

Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHi 540
Db 1900 TGCAAGAAAGGACTACATCACAGCCAAATACATCGAGAGGAGATACGCAAGAAAGACA 1959

Qy 540 sAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh 560
Db 1960 CGCGGATAACCGCGGAGGCTTTCAGCTCTTTGCGAGGCGCTCAAAAACGAGAGATATTTT 2019

Qy 560 eGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs 580
Db 2020 TGGATTGCTCCCAAGCTTATGCTGATGTGTGATCTTTACGGAAAAAATCCCACTGGCCAA 2079

Qy 580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600
Db 2080 CGGACATGAGCGGATGAAACGGCCCTCCACTTTGCAGTTCAGATCCGTGGATCAACACCTC 2139

Qy 600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy 620
Db 2140 TCTTCACATTTGATGACTTTTGTAGTTTCAGACAGTGGGNAACCTGGATTAACAGACAGGAA 2199

Qy 620 sGlySerThrAlaLeuHisTyrCysLysLeuThrAspAsnAlaGluCysLeuLysLeuLe 640
Db 2200 AGGCAGCACAGCCCTCTGCACTACTGCTGCTGACGCAATGTCGAGTGCCTCAAGTTGCT 2259

Qy 640 uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIl 660
Db 2260 CCTCGGGGGGAAAGCCCTCCATTCGAGATAGCAAAACGAGTCAGGAGAGACTCCCTGGACAT 2319
```

QY 660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 680  
Db 2320 TGCCAAAGCGCTCAAGACGAGCACTGTGAGAGCTGTGACCCCAAGCTTATCTGGAG 2379  
QY 680 gPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluLeuAspG1 700  
Db 2380 ATTAAATTCACGTTACGTTGATATGATGAGCACTATCCACGAGACCTTGGATGA 2439  
QY 700 uSerAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgp 720  
Db 2440 AAGTATGACGACATGGATGAGAAATTCAGCCAGTCC-CAACCGCGGGAAGACCGGC 2498  
QY 720 rolleSerPheTyrClnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerleuAlaA 740  
Db 2499 CCATCAGCTTACCAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCA 2558  
QY 740 rGAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerileLeuGlnA 760  
Db 2559 GAGATGCTGCAACCTTCCAGAGGAGAGAGAGGGCTTTTCATGCCAGCATCTTGCAGA 2618  
QY 760 snGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProS 780  
Db 2619 ATGAGACTTACGAGCCCTCTGAGTGGCAGCCACCTCCGCCAGCCTGCAGCCGCCA 2678  
QY 780 erThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS 800  
Db 2679 GCACCACAGCGCCCGCCGCTTCCACAGGAATGTTGGCAAAAGTTTCAGACAGCCTCCT 2738  
QY 800 erAlaAsnThrLeuTyrPheThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820  
Db 2739 CTGCTTAACACCTGTGGAAGACAACCTCTGAAGTGTGGAGCGGTGGAAGCCGCGAGCGAT 2798  
QY 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThra 840  
Db 2799 CTTGCTCAGATCCGCGACTCTCCATCCACGCTGCCCTCTTCGCGTGCATCTACCA 2858  
QY 840 snProLeuThrProThrProProProAlaLysThrProSerValMetGluAlaL 860  
Db 2859 ATCCCTGTACCCACCGCGCCGCCACCGTTCGCAAGACGCCACCGCTAATGGAAGCCT 2918  
QY 860 euSerGlnProSerLysProAlaProProGlyLysSerGlnLysArgProProLeuP 880  
Db 2919 TGAGCCAGCCGAGCAAGCTCCCGCGCTGGGATCTTCAGATCAGGCCCCCCTCTGC 2978  
QY 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899  
Db 2979 CCCACAGCCGCCAGCGCCCTCCCGCAGAGAGAGCTGCGCGGGGCT-GACAAGTCC 3037  
QY 900 ThrProLeuThrAsnLysGlyGlnProProArgGlyProValAspLeuSerAlaThrGluAla 919  
Db 3038 ACCCCTGTACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCT 3097  
QY 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939  
Db 3098 CTGGGTCTCTGTCCAATGTATGGTCTCGAGCGCCCTGCAACCCATGCTCTAGGAAGTCG 3157  
QY 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959  
Db 3158 CAGGCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAAC 3217  
QY 960 ProAspGluLeuThrPheSerGluGlyAspVallelleValAspGlyGluGluAspGln 979  
Db 3218 CCCGATGAGCTCACCTTCTCCAGGGGATGTGATCATCTGTCGAGCGGGGAGGAGCAG 3277  
QY 980 GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999  
Db 3278 GAGTGGTGGATTGGCCACATTGATGGAGATCTGTGTCGAAAGGCGCATTCCTCCGCTGTCA 3337  
QY 1000 PheValHisPheIleAlaAsp 1006  
Db 3338 TTTGTGCACCTTATCGCTGAC 3358

RESULT 2

US-10-104-047-314  
; Sequence 314, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 314  
; LENGTH: 3123  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-314  
Alignment Scores:  
Pred. No.: 1,17e-102 Length: 3123  
Score: 110.00 Matches: 165  
Percent Similarity: 98.80% Conservative: 0  
Best Local Similarity: 98.80% Mismatches: 1  
Query Match: 10.93% Indels: 2  
DB: 17 Gaps: 0  
US-09-914-042-1 (1-1006) x US-10-104-047-314 (1-3123)  
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
Db 272 CCCCTGACCCCGCCAGCGCCGCCACCGCTTGGCAAGACGCCAGCGTAATGGAAGCCTTG 331  
QY 861 SerGlnProSerLysProAlaProProGlyLysSerGlnLysArgProProLeuP 880  
Db 332 AGCCAGCGGAGCAGCGCTGCCCGCTGGGATCTCAGATCAGGCCCCCCTCTGCCCC 391  
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSerTh 900  
Db 392 CCACAGCGCCCGCCAGCGCTCCCGCAGAGAGCGCTGCGCGGGGCT-GACNAAGTCCAC 450  
QY 900 rProLeuThrAsnLysGlyGlnProProArgGlyProValAspLeuSerAlaThrGluAlaLe 920  
Db 451 CCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 510  
QY 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940  
Db 511 GGGTCTCTGTGCCAATGCTATGGTCTGCGAGCCCCCTGCAACCCATGCTTAGGAAGTCCGA 570  
QY 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960  
Db 571 GGCAACCAAGTTGAGCGCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACNAACCC 630  
QY 960 oAspGluLeuThrPheSerGluGlyAspVallelleValAspGlyGluGluAspGlnG1 980  
Db 631 CGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGCGAGGAGGAGGAGGAGGAGG 690  
QY 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
Db 691 GTGGTGGATTGGCCACATTGATGGAGATCTGTGGTTCGCAAGGCGCATTCCTCCGCTGTCA 750  
QY 1000 eValHisPheIleAlaAsp 1006  
Db 751 TGTGCACCTTATCGCTGAC 769  
RESULT 3  
US-09-764-875-37  
; Sequence 37, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PUZ02

```

; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 3436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-37

Alignment Scores:
Pred. No.: 1,28e-102 Length: 3436
Score: 110.00 Matches: 182
Percent Similarity: 98.91% Conservatve: 0
Best Local Similarity: 98.91% Mismatches: 1
Query Match: 10.93% Indels: 2
DB: 11 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-764-875-37 (1-3436)

Qy 612 GlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThr 631
Db 1 GGAACCTGGATAAACAGACAGGAAAGGAGCAGAGCCCTGCACCTACTGCTGCCTGACC 60
Qy 632 AspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsn 651
Db 61 GACAATGCCAGTGCTCAAGTTGCTCTCGGGGGAAGGCCCTCCATCGAGATAGCAAC 120
Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671
Db 121 GAGTCAGGAGAGACTCCGCTGGACATTGGCAAGCGCTCAAGCAGGAGCACTGTGAGGAG 180
Qy 672 LeuLeuThrClnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691
Db 181 CTGCTGACCAAGCCCTTATCTGGAAGATTAAATTCACGTTCAAGTTGAATATGAATGG 240
Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711
Db 241 CGACTACTCCAGAGAACCTGGATGAAGTAGATGACGACATGGATGAGAAATGGACGCC 300
Qy 712 SerGlu-AsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731
Db 301 AGTCC-CAACCGCGGGAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCACAGCT 359
Qy 731 uGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnAr 751
Db 360 TCAGTCTAACCGCTGTATCTTTGGCCACAGATGCTGCAACCTTGCCAGGAGACGAG 419
Qy 751 GAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPr 771
Db 420 GGCTTTTCATGCCAGACATCTTGCAAGATGAGACTTACGAGGCCCTCTGAGTGGCAGGCC 479
Qy 771 oProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAs 791
Db 480 ACCTCCGGCCAGCCTGCGAGCCCCCAGCACCACCGCCCCCCCCCTTCTCCACGGAA 539
Qy 791 nValGlyLys 794
Db 540 TGTGGCAAA 549

RESULT 4
US-09-764-875-358
; Sequence 358, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 3413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-38

Alignment Scores:
Pred. No.: 1,53e-70 Length: 3413
Score: 79.00 Matches: 157
Percent Similarity: 97.52% Conservatve: 0
Best Local Similarity: 97.52% Mismatches: 2
Query Match: 7.85% Indels: 4
DB: 11 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-764-875-358 (1-3413)

Qy 848 ProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAla 867
Db 572 CCACCCGTTGCCAAGACGCCCGCAGCGTAATGGAAGCCTTGAGCGCAGCGAGCCTGCC 631
Qy 868 ProProGlyLysSerGlnIleArgProProLeuProProGlnProProSerArgLeu 887
Db 632 CCGCCCTGGGATCTTCACAGATCAGGCCCCCACTCTGCCCCCAAGCGCCCGCCGCTC 691
Qy 888 ProGlnLysLysProAlaPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyCl 907
Db 692 CCGCAGAGAAGAGCGCTCGCGCGGGGCT-GACAAGTCCACCCCACTGACCAACAAGGCCA 750
Qy 907 nProArgGlyProValAspLysSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMe 927
Db 751 ACCGAGAGACCTGTGGATCTCTTGCAACGGAAGCTCTGGGCTCTCTGTCCAATGCTRT 810
Qy 927 t-ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProL 947
Db 811 -GGTCTCTGAGCCCTCGCACCCTAGCCCTAGGAAGTCGCGAGCAACCAAGTTGAAGCTTA 869
Qy 947 YsArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerG 967
Db 870 AGCGGGTGAAGCGCTCTATNACTGTGTGCTGACACACCCCGATGACCTCACCTTCTCG 929
Qy 967 luGlyAspValIlelleValaspGlyGluGluaspGlnGluTrpIleGlyHisIleAla 987
Db 930 AGGGGATGTGATCATCGTGGACGGGAGGAGCAGGAGTGGTGGATTGGCCACATTG 989
Qy 987 spGlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 990 ATGGAGATCTGTGTCAAAGGCGCATTCCTGGTGTCAATTTGTGCACTTTATCGCTGAC 1048

RESULT 5
US-09-918-995-27010
; Sequence 27010, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27010
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(442)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27010

```

## Alignment Scores:

Pred. No.: 2,53e-70 Length: 442  
 Score: 78.00 Matches: 78  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.75% Indels: 0  
 DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-918-995-27010 (1-442)

Qy 365 CysPheAspLeuIleSerHisAspArgThrTrpHisPheGlnAlaGluAspGlnGlu 384  
 Db TGCITTGACCTCAITTCATGACAGAACTTACCACTTCAAGCTGAAGATGAACAGGAA 97  
 Qy 385 CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPhe 404  
 Db TGTCAATATGGATGCTGTCTGCTCAAAATAGCAAGAAGAGCTTTAAACAATGCAITTT 157  
 Qy 405 LysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIle 424  
 Db AAGGGGATGACAAATACATCTGGAGAAATAACATCGTCCAAAGAACTGACAAAGGAGATCATC 217  
 Qy 425 SerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaPro 442  
 Db TCAGAAAGTCAGAGGATGACGGCAATGACGTCTGTGACTGTGGGGCGCCA 271

## RESULT 6

US-09-815-343-1154  
 ; Sequence 1154, Application US/09815343  
 ; Patent No. US20010055596A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meagher, Madeleine  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: King, Gordon E.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121.504  
 ; CURRENT APPLICATION NUMBER: US/09/815,343  
 ; CURRENT FILING DATE: 2001-03-22  
 ; NUMBER OF SEQ ID NOS: 1556  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1154  
 ; LENGTH: 563  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(563)  
 ; OTHER INFORMATION: n = A,T,C or G

## Alignment Scores:

Pred. No.: 3,53e-39 Length: 563  
 Score: 48.00 Matches: 131  
 Percent Similarity: 96.32% Conservative: 0  
 Best Local Similarity: 96.32% Mismatches: 2  
 Query Match: 4.77% Indels: 5  
 DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-815-343-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
 Db AACATGCTATTAAAGGGGATGACAAATCTGGAGAAATAACATCGTCCAAAGAACTGACA 61  
 Qy 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAsnAspValCysCysAspCysG1 440  
 Db AAGGAGATCATCTCAGAAGTGCT-GAGGATGACGGCAATGACGTCTGTGACTGTGG 120  
 Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysE 460  
 Db GCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTGACCTGACCTGAGTGTTC 180

Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspV 480  
 Db CGGAATCCACCGAGAGCTGGGGTTTCATTATTCAG-CATCGAGTCCCTGACCTTAGATG 239  
 Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500  
 Db TACTGGGAACATCTGAGCTGCTGCGCCAAAGAATATTGGGAATGACGGCTTTAATGAGA 299  
 Qy 500 leMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMet 519  
 Db TCATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGAGCCGACATG 359  
 Qy 520 AsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533  
 Db AATGCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGG 401

## RESULT 7

US-10-097-105-1154  
 ; Sequence 1154, Application US/10097105  
 ; Publication No. US20040037842A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meagher, Madeleine Joy  
 ; APPLICANT: King, Gordon E.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Xu, Jiangchun  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121.504C1  
 ; CURRENT APPLICATION NUMBER: US/10/097,105  
 ; CURRENT FILING DATE: 2002-03-13  
 ; NUMBER OF SEQ ID NOS: 1562  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1154  
 ; LENGTH: 563  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 403\_446, 464, 465, 472, 483, 485, 499, 539, 554  
 ; OTHER INFORMATION: n = A,T,C or G

## Alignment Scores:

Pred. No.: 3,53e-39 Length: 563  
 Score: 48.00 Matches: 131  
 Percent Similarity: 96.32% Conservative: 0  
 Best Local Similarity: 96.32% Mismatches: 2  
 Query Match: 4.77% Indels: 5  
 DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-097-105-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
 Db AACATGCTATTAAAGGGGATGACAAATCTGGAGAAATAACATCGTCCAAAGAACTGACA 61  
 Qy 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAsnAspValCysCysAspCysG1 440  
 Db AAGGAGATCATCTCAGAAGTGCT-GAGGATGACGGCAATGACGTCTGTGACTGTGG 120  
 Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysE 460  
 Db GCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTGACCTGACCTGAGTGTTC 180  
 Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspV 480  
 Db CGGAATCCACCGAGAGCTGGGGTTTCATTATTCAG-CATCGAGTCCCTGACCTTAGATG 239  
 Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500  
 Db TACTGGGAACATCTGAGCTGCTGCGCCAAAGAATATTGGGAATGACGGCTTTAATGAGA 299

```
Oy 500 leMetGluCysCysLeuProAlaGluAspSerValIysProAsnProClySer-AspMet 519
Db 300 TCATGGGAATGTGCTACCACTGAGGACTCAGTCAAAACCCAGCCAGCGACATG 359
Oy 520 AsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533
Db 360 AATCGAAGAGGACTACATCACAGCCAAAGTACATCGAGAGG 401

RESULT 8
US-10-027-632-34151
; Sequence 34151, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; POLYMORPHISMS: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34151
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34151

Alignment Scores:
Pred. No.: 1.29e-28 Length: 958
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.78% Indels: 0
DB: 13 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-027-632-34151 (1-958)

Oy 673 LeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArg 692
Db 41 CTGACCCAAAGCCTTATCTGGAGATTAAATCTCACGTTTCAGTTGAATATGAATGGCGA 100

Oy 693 LeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
Db 101 CTACTCCACGAAGACCTGGATGAAAGTGATGACGACATGGATGAGAAATTCGAG 154

RESULT 9
US-10-027-632-34151
; Sequence 34151, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; POLYMORPHISMS: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34151
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34151

Alignment Scores:
Pred. No.: 1.29e-28 Length: 958
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.78% Indels: 0
DB: 17 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-027-632-34151 (1-958)

Oy 673 LeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArg 692
Db 41 CTGACCCAAAGCCTTATCTGGAGATTAAATCTCACGTTTCAGTTGAATATGAATGGCGA 100

Oy 693 LeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
Db 101 CTACTCCACGAAGACCTGGATGAAAGTGATGACGACATGGATGAGAAATTCGAG 154

RESULT 10
US-10-276-774-404/c
; Sequence 404, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 404
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-404

Alignment Scores:
Pred. No.: 1.12e-20 Length: 409
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-276-774-404 (1-409)

Oy 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysProphe 146
```



; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22014  
; LENGTH: 60  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-975-22014

Alignment Scores:  
Pred. No.: 4.63e-10 Length: 60  
Score: 19.00 Matches: 19  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.89% Indels: 0  
DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-908-975-22014 (1-60)

Qy 525 TyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysHisAlaAspAen 543  
Db 3 TACATCACAGCCAAAGTACATCGAGGAGATACGCAAGGAAGACGCGGATAAC 59

## RESULT 14

US-09-815-343-1151  
; Sequence 1151, Application US/09815343  
; Patent No. US20010055596A1  
; GENERAL INFORMATION:

; APPLICANT: Meagher, Madeleine  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.504  
; CURRENT APPLICATION NUMBER: US/09/815,343  
; CURRENT FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1151  
; LENGTH: 326  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(326)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1151

Alignment Scores:  
Pred. No.: 2.58e-07 Length: 326  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-815-343-1151 (1-326)

Qy 411 GlyGluAenAenIleValGlnGluLeuThrLysGluIleIleSerGluVal 427  
Db 32 GGAGAAATAACATCGTCCAAAGAACTGACAAAGGAGATCATCTCAGAAAGTG 82

## RESULT 15

US-10-097-105-1151  
; Sequence 1151, Application US/10097105  
; Publication No. US20040037842A1  
; GENERAL INFORMATION:  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: King, Gordon E.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Harlocker, Susan L.

; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.504C1  
; CURRENT APPLICATION NUMBER: US/10/097,105  
; CURRENT FILING DATE: 2002-03-13  
; NUMBER OF SEQ ID NOS: 1562  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1151  
; LENGTH: 326  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 14, 29, 85, 105, 149, 150, 173, 208, 220, 223, 240, 252,  
; LOCATION: 256, 286, 285, 299, 302, 304, 318, 322  
; OTHER INFORMATION: n = A,T,C or G  
US-10-097-105-1151

Alignment Scores:  
Pred. No.: 2.58e-07 Length: 326  
Score: 17.00 Matches: 17  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-097-105-1151 (1-326)

Qy 411 GlyGluAenAenIleValGlnGluLeuThrLysGluIleIleSerGluVal 427  
Db 32 GGAGAAATAACATCGTCCAAAGAACTGACAAAGGAGATCATCTCAGAAAGTG 82

Search completed: August 5, 2005, 04:01:52  
Job time : 1466 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 17:15:00 ; Search time 393 Seconds  
(without alignments)  
4188.535 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDGPRKAGFPVSVFHIAD 1006

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2400006

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool\_b/US09914042/runat\_04082005\_090748\_2684/app.query.fasta\_1.1159  
-DB-Issued\_Patents=NA -QMT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042 @CGN 1 1 80 @runat\_04082005\_090748\_2684 -NCPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	29.5	4486	1	US-08-322-742-15
2	257	25.5	4328	1	US-08-322-742-12
3	68	6.8	2949	4	US-09-023-905A-8
4	68	6.8	5954	4	US-09-023-905A-6
5	32	3.2	2712	4	US-09-023-905A-11
6	32	3.2	4595	4	US-09-023-905A-9
7	30	3.0	3456	4	US-09-023-905A-5
8	30	3.0	4382	4	US-09-023-905A-3
9	30	3.0	5330	4	US-09-023-905A-1
10	16	1.6	213	4	US-09-513-999C-22310
11	12	1.2	4190	3	US-08-938-291A-2
12	12	1.2	4190	4	US-09-589-619-2

13	10	1.0	417	4	US-09-432-709A-140
c 14	10	1.0	697	4	US-09-432-709A-446
15	9	0.9	332	4	US-09-513-999C-384
16	9	0.9	514	4	US-09-854-133-660
17	9	0.9	601	4	US-09-949-016-150613
18	9	0.9	601	4	US-09-949-016-150614
19	9	0.9	601	4	US-09-949-016-150615
20	9	0.9	909	4	US-09-252-991A-12690
21	9	0.9	1131	3	US-09-383-586-7
22	9	0.9	1131	4	US-09-823-038A-7
23	9	0.9	1149	4	US-09-252-991A-12529
24	9	0.9	1295	1	US-08-433-854-3
25	9	0.9	1295	1	US-08-174-745A-3
26	9	0.9	1295	2	US-08-195-947-3
27	9	0.9	1295	2	US-08-433-885-3
28	9	0.9	1295	2	US-08-433-908B-3
29	9	0.9	1295	3	US-08-410-614-3
30	9	0.9	1640	4	US-09-573-080A-144
31	9	0.9	2119	4	US-09-620-312D-35
32	9	0.9	2366	4	US-09-270-767-13819
33	9	0.9	6990	4	US-09-620-312D-155
34	9	0.9	7158	4	US-09-949-016-14106
35	9	0.9	15378	3	US-08-785-420-1
36	9	0.9	37254	4	US-09-949-016-15973
c 37	9	0.9	38575	4	US-09-949-016-17304
38	9	0.9	38772	4	US-09-949-016-12732
39	9	0.9	38772	4	US-09-949-016-12729
40	9	0.9	77618	4	US-09-949-016-11768
41	9	0.9	94748	4	US-09-949-016-12648
c 42	9	0.9	94758	4	US-09-949-016-16741
c 43	9	0.9	108440	4	US-09-949-016-12065
c 44	9	0.9	108441	4	US-09-949-016-14090
45	9	0.9	113379	4	US-09-949-016-17561

#### ALIGNMENTS

#### RESULT 1

US-08-322-742-15  
; Sequence 15, Application US/08322742  
; Patent No. 5688641  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/322,742  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/938,823  
; FILING DATE: September 1, 1992  
; APPLICATION NUMBER: 07/844,296  
; FILING DATE: February 28, 1992  
; APPLICATION NUMBER: 07/552,216  
; FILING DATE: February 28, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00530/048003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070

Sequence 140, App  
Sequence 446, App  
Sequence 384, App  
Sequence 660, App  
Sequence 150613,  
Sequence 150614,  
Sequence 150615,  
Sequence 12690, A  
Sequence 7, Appli  
Sequence 12529, A  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 144, App  
Sequence 35, Appli  
Sequence 13819, A  
Sequence 155, A  
Sequence 14106, A  
Sequence 1, Appli  
Sequence 15973, A  
Sequence 17304, A  
Sequence 12382, A  
Sequence 12729, A  
Sequence 11768, A  
Sequence 12648, A  
Sequence 16741, A  
Sequence 12065, A  
Sequence 14090, A  
Sequence 17561, A

; TELEFAX: (617) 542-8506  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4486  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 US-08-322-742-15

## Alignment Scores:

Pred. No.: 6, 46e-289 Length: 4486  
 Score: 297.00 Matches: 701  
 Percent Similarity: 98.59% Conservative: 0  
 Best Local Similarity: 98.59% Mismatches: 4  
 Query Match: 29.52% Indels: 10  
 DB: 1 Gaps: 0

US-09-914-042-1 (1-1006) x US-08-322-742-15 (1-4486)

Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLeuArgLys 321  
 Db 1 AAGGAACATGGACCGCGAAGCGGAGCCCTTACAAAGAGAGTGACGGGATCCGAAAA 60  
 Qy 322 ValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341  
 Db 61 GTGTGGCAGAAAGGAAATGTTTCAGTTAAATAATGTTTCTGACCATTATCCCATGGTACC 120  
 Qy 342 AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProG1 361  
 Db 121 GCTAACCGGGCTCTCTGCAAGCTCAACTGCTTAACCTGCCAGGTGAAGACCAACCTGA 180  
 Qy 361 uGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAs 381  
 Db 181 GGAGAAGAAGTGCTTTTGACCTTATTTTCATGACAGAACTTACCACCTTCAAGCTGAAGA 240  
 Qy 381 pGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluAlaLeuAs 401  
 Db 241 TGAACAGGAATGTCAAAATGATGATGCTGTGTGCTGCAAAATAGCAAGAAGAGCTTTAAA 300  
 Qy 401 nAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys 421  
 Db 301 CAATGCATTTAAGGGGGATGACAACTACTGGAGAAATAACATCGTCCAAAGACTGACAAA 360  
 Qy 421 sGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspGlyAl 441  
 Db 361 GGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGGGCG 420  
 Qy 441 aProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerG1 461  
 Db 421 GCCAGATCTTACATGGCTTTCCACCAACCTGGGCATCTCTGACCTGCATCGAGTGTTCGG 480  
 Qy 461 VileHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValL 481  
 Db 481 AATCACCAGAGAGCTGGGGTTCATTATTCCAG-GATGCAGTCCCTGACCTAGATGTAC 539  
 Qy 481 euGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleM 501  
 Db 540 TGGGAACATCTGAGCTGTCTCGCCCAAGAAATTTGGGAATTCAGGCTTTTATGAGATCA 599  
 Qy 501 etGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnA 521  
 Db 600 TGGAAATGTTGCTTACAGCTGAGGACTCAGTCAAAACCCCAACCCAGCGAGCATGAATG 659  
 Qy 521 laArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisA 541  
 Db 660 CAAGAAGAGGACTATCATCAGCCCAAGTACATCAGAGGAGATACCAAGGAAGAGCAGC 719  
 Qy 541 laAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheG 561  
 Db 720 CGGATAACCGCGGAGGCTTTCAGCTTTTGGAGGCCGTCAAAACGAGAGATATTTTGT 779  
 Qy 561 lYLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnG 581

Db 780 GATTGCTCCAAGCTTATGCTGTGGATCTTACGGAATAATCCCACTGCCCAACG 839  
 Qy 581 lYHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerL 601  
 Db 840 GACATGAGCCGGATGAAACGGCCCTCCACCTTGCACTCAGATCGATCGATCGAACCTCTC 899  
 Qy 601 euHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysG 621  
 Db 900 TTCACATTTGTAGACTTTTATGTTTCCAGAACAGTGGGAACCTGGATATAACAGACGGGAAG 959  
 Qy 621 lYSerThrAlaLeuHisTyrCysValLeuThrAspAsnAlaGluCysLeuLysLeuLeuL 641  
 Db 960 GCAGACAGCCCTGCAGCTACTGCTGCTGACCGCAATGCGAGTGCCTCAAGTTGCTCC 1019  
 Qy 641 euArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleA 661  
 Db 1020 TCGGGGGAGGCGCTCCATCGAGATAGCAATAGTCCAGAGAGACTCCGCTGGACATTG 1079  
 Qy 661 laLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgP 681  
 Db 1080 CCAAGCCCTCAAGCAGCAGCACTGTGAGGAGCTGTGACCCAGACCTTATCTGGAAGAT 1139  
 Qy 681 heAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluS 701  
 Db 1140 TTAATTTCTCAGTTCACTGTTGAATATGAATGGCCACTACTCCACGAGACCTGGTGA 1199  
 Qy 701 erAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgPro 720  
 Db 1200 GTGATGACGACATGGATGAGAAATTTGACGCCAGTCC-CAACCCGCGGGAAGACCGGCC 1258  
 Qy 721 lIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
 Db 1259 ATCAGCTTCTACCAAGCTGGGCTCCAAACAGCTTCACTTAACGCTGTATCTTTGGCC 1318  
 Qy 741 AspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuGlnAs 760  
 Db 1319 GATGCTGCAAACTTGGCCNAGGA-CAAGCAGAGGGCTTTTCATGCCAGCATCTTGCAG 1377  
 Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAProSe 780  
 Db 1378 TGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCCACTCCCGCCAGCTGCAGGCCCCAG 1437  
 Qy 780 rThrThrSerAla-ProProLeuProProArgAsnValGlyValGlnThrAlaSerS 800  
 Db 1438 CACCACAGCGCCCGCCCGCTTCTCCACGGAATTTGGCAAGTTTCAGACAGCTCTCT 1497  
 Qy 800 erAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820  
 Db 1498 CTGCTAACACCTGTGGNAGACAACTCTGTAAGTGTGGACGGTGGAGCGCGCAGCGAT 1557  
 Qy 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA 840  
 Db 1558 CTTCGTTCAGATCCGCACTGTCCATCCACGCTGCCCGCTCTTCGCGTGACATCTACCA 1617  
 Qy 840 snProLeuThrProThrProProValAlaLysThrProSerValMetGluAlaL 860  
 Db 1618 ATCCCTTACCCCGCCCGCCCGCTTCCAGAGACCGCCCGCTTATGGAAGCT 1677  
 Qy 860 euSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuP 880  
 Db 1678 TGAGCCAGCCGAGCAAGCTGCCCGCTGGGATCTCAGATCAGGCCCGCCCGCTCTGCG 1737  
 Qy 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899  
 Db 1738 CCCCACAGCCCGCCCGCTCCCGCAGAGAAGAGCTTCGCGCAGGGGCT-GACAAGTCC 1796  
 Qy 900 ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla 919  
 Db 1797 ACCCCACTGACCAAAAGGCCCAACCGAGAGAGCTGTGGATCTCTCTGCAACCGAGACT 1856  
 Qy 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939

Db 1857 CTGGTCTCTGTCTCAATGCTATGTCTCTGAGCCCTGACCCCTGACCCATGCTAGGAGTCG 1916  
Qy 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959  
Db 1917 CAGGCAACCAAGTTGAAGCCTAAGCGGTGAAAGCGCTCTATAACTGTGTGGCTGACAAC 1976  
Qy 960 ProAspGluLeuThrPheSerGluClyAspValIleIleValAspGlyGluGluAspGln 979  
Db 1977 CCGGATGAGCTCACCTCTCCGAGGGGATGTGATCATCGTGAGCGGGAGGAGACCAG 2036  
Qy 980 GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999  
Db 2037 GAGTGTGGATGGCCACATGATGAGATCTCTGTCTGCAAGGCCCAATCCCGGTGTCA 2096  
Qy 1000 PheValHisPheIleAlaAsp 1006  
Db 2097 TTTGTGCATTTATCGCTGAC 2117

## RESULT 2

US-08-322-742-12  
; Sequence 12, Application US/08322742  
; Patent No. 5688641  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/322,742  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/938,823  
; FILING DATE: September 1, 1992  
; APPLICATION NUMBER: 07/844,296  
; FILING DATE: February 28, 1992  
; APPLICATION NUMBER: 07/552,216  
; FILING DATE: February 28, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00530/048003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4328  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-322-742-12

Alignment Scores:			
Pred. No.:	1,35e-248	Length:	4328
Score:	257.00	Matches:	648
Percent Similarity:	98.63%	Conservative:	0
Best Local Similarity:	98.63%	Mismatches:	4
Query Match:	25.55%	Indels:	9
DB:	1	Gaps:	0

US-09-914-042-1 (1-1006) x US-08-322-742-12 (1-4328)

Qy 355 GlnVallyThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374  
Db 3 CAGGTGAAGACCAACCCCTGAGGAGAAGTGTCTTACCTTTATTTACATGACAGAACT 62  
Qy 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsn 394  
Db 63 TACCACCTTTCAAGCTGAAGATGAAACAGAAATGTCAAAATATGGATGTCTGTCTGCAAAAT 122  
Qy 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn 414  
Db 123 AGCAAGAGAGAGCTTTAAACAATGCATTTAAGGGGATGACATATCTGGAGAAAATTAAC 182  
Qy 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp 434  
Db 183 ATCGTCCAAGAACTGACAAAGAGATCATCTCAGAAGTGCAGAGATGACGGCAATGAC 242  
Qy 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu 454  
Db 243 GTCTGTGTGACTGTGGGGGCCAGATCTCATATGGCTTTCCACCAACCTGGGCATCTCG 302  
Qy 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1 474  
Db 303 ACCTGCATCGAGTGTTCGGGATCCACGAGAGCTGGGGTTCATTATTTCCAG-CATGCA 361  
Qy 474 nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs 494  
Db 362 GTCCCTGACCTTAGATGTACTGGGAACATCTGAGCTGCTGCTCGCCAAGATATTTGGGAA 421  
Qy 494 nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs 514  
Db 422 TGCAGGCTTTAATGAGATCATGGAATGTGCTTACCAGCTGAGGACTCAGTCAAAACCCAA 481  
Qy 514 nProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgAr 534  
Db 482 CCCAGGCGAGCATGATGCAAGAAGAGACTACATCACGCCAAGTACATCGAGAGAG 541  
Qy 534 gTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVa 554  
Db 542 ATACGCAAGGAAGAAGCAGCGGATACCGCGCGAAGCTTCACAGCTTTTCGAGGCGCT 601  
Qy 554 lLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGl 574  
Db 602 CAAAACGAGAGATATTTTGGATTGCTCCAAAGCTTATGCTGATGTTGGATCTTTACGGA 661  
Qy 574 uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr 594  
Db 662 AAAATCCCACTGGCCCAACGAGCATGAGCCGATGAACGGCCCTCCACCTTTGCAGTCAG 721  
Qy 594 gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe 614  
Db 722 ATCGTGGATCGAACCCTCTCTTCACATTTGTAGACTTTTTCAGAACAGTGGGAACCT 781  
Qy 614 uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAl 634  
Db 782 GGATAAAACAGACAGGGAAGGAGCAGCAGCGCCCTGCACCTACTGCTGCTGAGCCGACAATGC 841  
Qy 634 aGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGl 654  
Db 842 CGAGTGGCTCAAGTTGCTCTCTCGGGGGAAGGCCCTCCATCGAGATAGCAATAGTCTAGG 901  
Qy 654 yGluThrProLeuAspIleAlaLysArgLysGlyHisGluHisCysGluGluLeuLeuTh 674  
Db 902 AGAGACTCCGCTGGACATTTGCCAAGCGCTCAACGACGAGCAGTGTGAGGAGCTGCTGAC 961  
Qy 674 rGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLe 694  
Db 962 CCAAGCCTTATCTGGAAGATTTAATTTCTCAGTTTCAAGTTGAATATGAATGGCAGTACT 1021  
Qy 694 uHisGluAspLeuAspGluSerAspAspMetAspGlyLysLeuGlnProSerGlu-A 714  
Db 1022 CCAGAAAGACCTTGATGAAAGTGTATGACGATGGATGAGAAATTTGCAGGCCCGCTCC-CA 1080

```
Qy 714 snArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerA 734
Db 1081 ACCGGCGGAGAGACCGGCCCATCAGCTTCTACAGCTGGCTCCACAGCTTTCAGTCTA 1140
Qy 734 snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe 753
Db 1141 ACGTGATATCTTTGGCCAGAGATGCTGCAACCTTGCCAAAGGA-CAAGCAGAGGGCTTTC 1199
Qy 754 MetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProPro 773
Db 1200 ATGCCACAGCTCTTGCAATGAGACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCC 1259
Qy 774 AlaGlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValG1 793
Db 1260 GCCAGCCTTGAGCCCGCCAGCACCCAGCGGCCCGCCCGCTTCTCCACGGAATGTGG 1319
Qy 793 YLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValAs 813
Db 1320 CAAAGTTTCAGACAGCCTCTCTGCTAACACCTGTGGAAGACAAACTCTGTAAGTGTGA 1379
Qy 813 pGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProPr 833
Db 1380 CGGTGGAAGCGGCGGCGATCTTCGTAGATCCGCGCAGCTGTCCATCCACGCTGCCGCC 1439
Qy 833 oLeuArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysTh 853
Db 1440 TCTTCGGGTGACATCTCAATCCCTGACCCCGCCAGCGGCCCGCCCGCTTGCCAGAC 1499
Qy 853 rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerG1 873
Db 1500 GCCAGCGCTAATGGAAGCTTGAGCCAGCCGAGCAAGCTGCCCGCGCTGGGATCTACA 1559
Qy 873 nileArgProProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAl 893
Db 1560 GATCAGGCCCGCCACCTCTGCCCCCAGCCGCGCCAGCGCCCTCCCGCAGAGAAGCCCTGC 1619
Qy 893 apro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA 913
Db 1620 GCCAGGGGCT-GACNAGTCCACCCCACTGACCAACAAAGGCCAACCCGAGAGAGACTGTGG 1678
Qy 913 spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA 933
Db 1679 ATCTCTGTCAAGGGAAGCTCTGGGTCTCTGTCTCAATGCTATGCTCTCGAGCCCGCTG 1738
Qy 933 laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT 953
Db 1739 CACCCATCCCTAGGAAGTCGAGGCAACCAAGTTGAAGCCTTAAGCGGGTGAAGCGCTCT 1798
Qy 953 YrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleV 973
Db 1799 ATAACTGTGTGCTGACAAACCCGATGAGCTCACCTTCTCCAGGGGGATGTGATCATCG 1858
Qy 973 alAspGlyGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgL 993
Db 1859 TGGACGGGGAGAGGAGCAGGAGTGGTGGATTGGCCACATTGATGGAGATCTCTGTGTGCA 1918
Qy 993 ysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 1919 AAGCGCATCCCGGTGTCAATTTGTGCATTTATCGCTGAC 1959
```

## RESULT 3

```
US-09-023-905A-8
; Sequence 8, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
```

```
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2949
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-8

Alignment Scores: 3.53e-58 Length: 2949
Pred. No.: 68.00 Matches: 68
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 6.76% Indels: 0
Query Match: 4 Gaps: 0
DB:

US-09-914-042-1 (1-1006) x US-09-023-905A-8 (1-2949)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGlu 190
Db 511 CACGGATGATCCGAGCGGAGATCAGCGGAGCAGATAGCAGAGAGATGGAAGAGAG 570
Qy 191 ArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIle 210
Db 571 CGCGCTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAGTCAATGAATCAAGATC 630
Qy 211 LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe 230
Db 631 AAAAAAGGTGTGACCTGTCTCCAGAACTCATCAATACTTCCACGCACATGTCACACTTC 690
Qy 231 PheGlnAspGlyLeuLysAlaVal 238
Db 691 TTTCAGGATGGTCTCAAGCGGTG 714

RESULT 4
US-09-023-905A-6
; Sequence 6, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 5954
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (433) ... (3378)
US-09-023-905A-6

Alignment Scores: 7.07e-58 Length: 5954
Pred. No.: 68.00 Matches: 68
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00%
```

```
Query Match: 6.76% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-6 (1-5954)

Qy 171 HisGlyMetIleAqThrGluIleSerGlyAlaGluIleAlaGluGluMetGluIleGlu 190
Db 943 CACGGAATATCCGGACCGAGATACGCGAGCAGATAGCAGAGATGGAAGAG 1002
Qy 191 ArgArgPheGlnLeuMetCysGluTyrLeuLeuLeuValAsnGluIleIle 210
Db 1003 CGGGTTTCTCCAGCTTCAGATGTGTCAGTACCTCTCAAAGTCAATGAATCAAGATC 1062
Qy 211 LysIleGlyValAspLeuLeuGlnAsnLeuIleIleTyrPheHisAlaGlnCysAsnPhe 230
Db 1063 AAAAAAGGTGCGACCTCTCAGAAATCTCATCAATAATCTTCCAGCGCACAGTGCACACTTC 1122
Qy 231 PheGlnAspGlyLeuLeuAlaVal 238
Db 1123 TTTCAGGATGGTCTCAAAGCGGTG 1146

RESULT 5
US-09-023-905A-11
; Sequence 11, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-11
Alignment Scores:
Pred. No.: 6.51e-22 Length: 2712
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.18% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-11 (1-2712)

Qy 125 PheProLeuAspSerLeuLeuIleGlyAspLeuIleGlyValIleGlyAspLeuIle 144
Db 373 TTTCACCTAGACAGTCTGCTTAAGGAGACCTCAAAGGAGTCAAAGGGGATTTGAAAAAG 432
Qy 145 PropheAspLysAlaTrpLysAspTyrGluThrLys 156
Db 433 CCATTGTATAAGCATGGAAGGATTTGAAACCAA 468

RESULT 6
US-09-023-905A-9
; Sequence 9, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 4595
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-9
Alignment Scores:
Pred. No.: 1.1e-21 Length: 4595
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.18% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-9 (1-4595)

Qy 125 PheProLeuAspSerLeuLeuIleGlyAspLeuIleGlyValIleGlyAspLeuIle 144
Db 672 TTTCACCTAGACAGTCTGCTTAAGGAGACCTCAAAGGAGTCAAAGGGGATTTGAAAAAG 731
Qy 145 PropheAspLysAlaTrpLysAspTyrGluThrLys 156
Db 732 CCATTGTATAAGCATGGAAGGATTTGAAACCAA 767

RESULT 7
US-09-023-905A-5
; Sequence 5, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-5
Alignment Scores:
Pred. No.: 8.6e-20 Length: 3456
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 4 Gaps: 0
```

US-09-914-042-1 (1-1006) x US-09-023-905A-5 (1-3456)

Qy 127 LeuAspSerLeuLeuGlyAspLeuLysGlyVallysGlyAspLeuLysLysProphe 146  
Db 439 CTGGACTCTCTGCTCAAGAGGAGATCTAAAGGAGTGAAGGGGACCTTTAAAAAGCCTTTC 498  
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156  
Db 499 GACAAGGCTGGAAGAGACTATGAAACCAAG 528

## RESULT 8

US-09-023-905A-3  
; Sequence 3, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 4382  
; TYPE: DNA  
; ORGANISM: Danio rerio  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (351)...(3803)  
US-09-023-905A-3

Alignment Scores:  
Pred. No.: 1.09e-19 Length: 4382  
Score: 30.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.98% Indels: 0  
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-3 (1-4382)

Qy 127 LeuAspSerLeuLeuGlyAspLeuLysGlyVallysGlyAspLeuLysLysProphe 146  
Db 789 CTGGACTCTCTGCTCAAGAGGAGATCTAAAGGAGTGAAGGGGACCTTTAAAAAGCCTTTC 848  
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156  
Db 849 GACAAGGCTGGAAGAGACTATGAAACCAAG 878

## RESULT 9

US-09-023-905A-1  
; Sequence 1, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5330  
; TYPE: DNA  
; ORGANISM: Bos sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (209)...(3595)  
US-09-023-905A-1

Alignment Scores:  
Pred. No.: 1.32e-19 Length: 5330  
Score: 30.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.98% Indels: 0  
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-1 (1-5330)

Qy 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyVallysGlyAspLeuLysLysProphe 146  
Db 647 TTGGATTCTCTTGTGAAAGGAGACCTGAAGGGAGTCAAGGCGATCTCAAGAAACCATTT 706  
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156  
Db 707 GACAAGCTTGGAAAGATTATGAGACGAAG 736

## RESULT 10

US-09-513-999C-22310  
; Sequence 22310, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 22310  
; LENGTH: 213  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 33  
; OTHER INFORMATION: v=a or c or g  
US-09-513-999C-22310

Alignment Scores:  
Pred. No.: 7.18e-07 Length: 213  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.59% Indels: 0  
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-513-999C-22310 (1-213)

Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGlu 465  
Db 36 AACTTGGGTATTTTGACCTGTATAGAAATGTTCTGGCATCCATAGGAA 83

## RESULT 11

```
US-08-938-291A-2
; Sequence 2, Application US/08938291A
; Patent No. 6117673
; GENERAL INFORMATION:
; APPLICANT: Lev, Sima
; APPLICANT: Plowman, Gregory D.
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: RDB PROTEINS AND RELATED
; TITLE OF INVENTION: PRODUCTS AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,291A
; FILING DATE: September 26, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,337
; FILING DATE: October 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 228/172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-938-291A-2
Alignment Scores:
Pred. No.: 0.147 Length: 4190
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.19% Indels: 0
DB: 3 Gaps: 0
US-09-914-042-1 (1-1006) x US-08-938-291A-2 (1-4190)
Qy 772 ProProAlaGlnProAlaAlaProSerThrThrSer 783
Db 2607 CCCCCGCGCCAGCCAGCGCCGCCAGCCACCAGT 2642
RESULT 12
US-09-589-619-2
; Sequence 2, Application US/09589619
; Patent No. 6576442
; GENERAL INFORMATION:
; APPLICANT: Lev, Sima
; APPLICANT: Plowman, Gregory D.
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: RDB PROTEINS AND RELATED
; TITLE OF INVENTION: PRODUCTS AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/589,619
; FILING DATE: 07-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,291
; FILING DATE: September 26, 1997
; APPLICATION NUMBER: 60/027,337
; FILING DATE: October 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 228/172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-589-619-2
Alignment Scores:
Pred. No.: 0.147 Length: 4190
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.19% Indels: 0
DB: 4 Gaps: 0
US-09-914-042-1 (1-1006) x US-09-589-619-2 (1-4190)
Qy 772 ProProAlaGlnProAlaAlaProSerThrThrSer 783
Db 2607 CCCCCGCGCCAGCCAGCGCCGCCAGCCACCAGT 2642
RESULT 13
US-09-492-709A-140
; Sequence 140, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
```

; SEQ ID NO 140  
; LENGTH: 417  
; TYPE: DNA  
; ORGANISM: E. Coli  
US-09-492-709A-140

## Alignment Scores:

Pred. No.:	1.57	Length:	417
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.99%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-492-709A-140 (1-417)

Qy 71 AenGluGluGlnTyrThrGlnAlaLeuGlu 80  
|||||  
Db 76 AACGAGGAACAATACACCCAGCGCTGGAA 105  
|||||

## RESULT 14

US-09-492-709A-446/c  
; Sequence 446, Application US/09492709A  
; Patent No. 6720139

## GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; FILE REFERENCE: ESCHERICHIA COLI

; CURRENT APPLICATION NUMBER: US/09/492,709A

; CURRENT FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 446

; LENGTH: 697

; TYPE: DNA

; ORGANISM: Escherichia coli

US-09-492-709A-446

## Alignment Scores:

Pred. No.:	2.6	Length:	697
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.99%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-492-709A-446 (1-697)

Qy 71 AenGluGluGlnTyrThrGlnAlaLeuGlu 80  
|||||  
Db 241 AACGAGGAACAATACACCCAGCGCTGGAA 212  
|||||

## RESULT 15

US-09-513-999C-384

; Sequence 384, Application US/09513999C

; Patent No. 6783961

## GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 384  
; LENGTH: 332

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 89..331

; FEATURE:

; NAME/KEY: sig\_peptide

; LOCATION: 89..208

; OTHER INFORMATION: score 4.4

; OTHER INFORMATION: seq LNLGLVLMCIKCSG/IH

US-09-513-999C-384

## Alignment Scores:

Pred. No.:	12.7	Length:	332
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.89%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-513-999C-384 (1-332)

Qy 456 CysleGluCysSerGlyIleHisArg 464  
|||||  
Db 191 TGTATTGAATGTTTCAGGAATCCACCGC 217  
|||||

Search completed: August 5, 2005, 00:25:15

Job time : 435 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 13:57:13 ; Search time 1228 Seconds  
(without alignments)

4849.560 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAFVSVFHFAD 1006

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8776198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTC spool/h/US09914042/runat\_04082005\_090746\_2651/app.query.fasta\_1.1159  
-DB=N Geneseq -QFMT=fastop -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042 @CGN 1.1 615 @runat\_04082005\_090746\_2651 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq 16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	25.5	4328	2	AQ28758 Partial s
2	110	10.9	3123	10	ADB62160 Human cDN
3	110	10.9	3436	4	ABK43447 DNA encod
4	110	10.9	3436	12	ADI53834 cDNA enco
5	79	7.9	3413	4	ABK43768 DNA encod

6	79	7.9	3413	12	ADI54155	Adi54155 cDNA enco
7	78	7.8	442	9	ACH39798	Ach39798 Human foe
8	68	6.8	5954	2	AAV59105	AAV59105 Zebrafish
9	48	4.8	563	4	AA558478	AA558478 CDNA #115
10	32	3.2	4595	2	AAV59106	AAV59106 Zebrafish
11	30	3.0	409	4	ABA08628	ABA08628 Human GTP
12	30	3.0	3775	10	ADD71214	Add71214 Human int
13	30	3.0	3946	10	ADD71207	Add71207 Human int
14	30	3.0	3974	12	ADJ51008	Adj51008 Human cDN
15	30	3.0	4382	2	AAV59104	AAV59104 Zebrafish
16	30	3.0	5330	2	AAV59103	AAV59103 Bovine di
17	30	3.0	5947	13	ADS34277	Ads34277 POSH prot
18	24	2.4	544	13	ADQ79413	Adq79413 Novel can
19	22	2.2	4870	13	ADS34272	Ads34272 POSH prot
20	22	2.2	4997	13	ADS10088	Ads10088 Human the
21	22	2.2	5475	13	ADS34273	Ads34273 POSH prot
22	21	2.1	345	4	AAI84720	Aai84720 Human pol
23	19	1.9	60	6	ABN49266	Abn49266 Human spl
24	17	1.7	326	4	AA558475	AA558475 CDNA #115
25	17	1.7	747	6	ABK71808	Abk71808 Human dit
26	16	1.6	213	3	AAAC18235	Aac18235 Human sec
27	16	1.6	469	4	ABA08512	AbA08512 Human dif
28	14	1.4	1038	4	ABA08863	AbA08863 Human sec
29	14	1.4	5033	4	AAI59224	Aai59224 Human pol
30	13	1.3	167	4	AAI46525	Aai46525 Probe #15
31	13	1.3	167	4	AKA40054	Aka40054 Human bon
32	13	1.3	167	4	ABSA0054	AbA0054 Human liv
33	13	1.3	167	5	AAI06958	Aai06958 Probe #69
34	13	1.3	323	6	ABQ97783	Abq97783 Mouse ES
35	13	1.3	446	4	AAI33385	Aai33385 Probe #20
36	13	1.3	446	4	AAK27473	Aak27473 Human bon
37	13	1.3	446	4	ABS27037	AbS27037 Human liv
38	13	1.3	446	5	AAI01973	Aai01973 Probe #19
39	12	1.2	296	6	ABL93107	AbI93107 Rat metas
40	12	1.2	296	12	ADN07897	Adn07897 Human mam
41	12	1.2	2503	3	AAAC69461	Aac69461 Human sec
42	12	1.2	3580	12	ADQ86427	Adq86427 Human tum
43	12	1.2	3580	13	ADQ85284	Adq85284 Human tum
44	12	1.2	3580	13	ADQ83337	Adq83337 Human tum
45	12	1.2	3710	13	ADR06841	Adr06841 Full leng

#### ALIGNMENTS

RESULT 1  
AAQ28758  
ID AAQ28758 standard; DNA; 4328 BP.  
XX AC AAQ28758;  
XX AC  
XX 25-MAR-2003 (revised)  
DT 25-FEB-1993 (first entry)  
XX  
XX Partial sequence of tumour suppressor gene U10.  
DE  
XX Can19; tumour suppressor gene; cancer; therapy; ss.  
XX Homo sapiens.  
XX  
XX WO9215602-A1.  
XX  
XX 17-SEP-1992.  
XX  
XX 28-FEB-1992; 92WO-US001624.  
XX  
XX 28-FEB-1991; 91US-00662216.  
XX  
XX (DAND ) DANA FARBER CANCER INST INC.  
XX  
XX PI Sager R;  
XX  
XX WPI; 1992-331663/40.  
XX

PT Diagnosis and treatment of cancer - using candidate tumour suppressor  
 XX genes or the corresp. antibodies.  
 PS Claim 30; Page 38-43; 54pp; English.

XX An adaptation of the subtractive hybridization technique was used which  
 CC utilizes a biotinylation-based subtraction procedure instead of  
 CC hydroxyapatite as previously used. In this procedure, a single strand  
 CC phagemid cDNA library from normal cell polyA+ mRNA is hybridized with  
 CC excess biotinylated tumor polyA+ mRNA, and the resulting double stranded  
 CC sequences are removed by binding to streptavidin. The remaining single-  
 CC stranded phagemid cDNAs are converted to double-stranded form and used to  
 CC transform bacterial host cells. The resulting subtracted cDNA library is  
 CC differentially screened with total cDNA from normal and tumor cells. This  
 CC method produced some 20 additional cloned cDNAs. Also found by this  
 CC method were several genes which, on the basis of the partial DNA  
 CC sequences appear to be novel sequences not previously entered into  
 CC GENBANK. The portion of the cDNAs so sequenced represents part of the  
 CC coding region and/or part of the 3' untranslated region of each cDNA (see  
 CC Q28749-58). (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-  
 CC MAR-2003 to correct PA field.)

SQ Sequence 4328 BP; 1236 A; 970 C; 912 G; 1210 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	1.74e-243	Length:	4328
Scores:	257.00	Matches:	648
Percent Similarity:	98.63%	Conservative:	0
Best Local Similarity:	98.63%	Mismatches:	4
Query Match:	25.55%	Indels:	9
DB:	2	Gaps:	0

US-09-914-042-1 (1-1006) x AAQ28758 (1-4328)

Qy	355	GlnValLysThrAsnProGluLysLysCysPheAspLeuLeuSerHisAspArgThr	374
Db	3	CAGGTGAAGACCAACCCCTGAGGAGAAAGTCTTGGACCTTATTCATGACAGAACT	62
Qy	375	TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrrMetSerValLeuGlnAsn	394
Db	63	TACCACCTTTCAGCTGAAGATGAACAGGAATGTCAAAATATGGATGTCTGTCTGCAAAAT	122
Qy	395	SerLysGluGluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn	414
Db	123	AGCAAGAAAGAAAGCTTTAAACATGCAATTAAGGGGGATGACATCTGGAGAAATTAAC	182
Qy	415	IleValGlnGluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAsp	434
Db	183	ATCGTCCAAAGAACTGACAAAGGAGATCATCTCAGAAAGTGACAGGATGACGGCAATGAC	242
Qy	435	ValCysCysAspCysGlyValaProAspProThrTrrLeuSerThrAsnLeuGlyIleLeu	454
Db	243	GTCTGCTGTGACTGTGGGGCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTGT	302
Qy	455	ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1	474
Db	303	ACCTGCATCGAGTGTTCGGGAATCACCAGAGAGCTGGGGGTTCATTATTCAG-CATGCA	361
Qy	474	nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs	494
Db	362	GTCCCTGACCTTAGATGTACTGGGAACATCTGAGCTGTCTGCTCGCAAGAAATATTGGAA	421
Qy	494	nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs	514
Db	422	TGCAGGCTTTAATGAGATCATGGAATGTGGCTTACCAGTGGAGTCTCAGTCAACCCAA	481
Qy	514	nProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgAr	534
Db	482	CCGAGCAGGCACATGATGCAAGAAAGGACTTACATCACGCCAGTATCATCGAGAGAG	541
Qy	534	gTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVa	554
Db	542	ATACGAGGAGGAGACCGCGGATAACCGCGGAAAGCTTTCACAGTCTTTTGCAGGCGGT	601

Qy	554	lLysThrArgAspIlePheGlyLeuLeuGlnAlaTrrAlaAspGlyValaAspLeuThrGl	574
Db	602	CAAAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGTGTGGATCTTACGGA	661
Qy	574	uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr	594
Db	662	AAAAATCCCACTGGCCCAACGACATGATGCGGATGAACGGCCCTCCACCTTGCAGTCAG	721
Qy	594	gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe	614
Db	722	ATCCGTGGATCGAACCTCTCTTCACATTTAGACTTTTATGTTTCAGAACACAGTGGGAACCT	781
Qy	614	uAspLysGluThrGlyLysGlySerThrAlaLeuHisTrrCysCysLeuThrAspAsnAl	634
Db	782	GGATAAACACAGACGGAAGGACGACAGCCCTGCATCTACTGCTGCTGACCCGCAATGC	841
Qy	634	gGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGl	654
Db	842	CGAGTGCCTCAAGTTGCTCTCTCGGGGGAAGGCCCTCCATCGAGATAGCAAAATGAGTCAGG	901
Qy	654	yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuTh	674
Db	902	AGAGACTCCGCTGGACATTCCAAGCGCCTCAAGCACGACGACCTGTGAGGAGCTGTGCAC	961
Qy	674	rGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTrrGluTrrArgLeuLe	694
Db	962	CCAAAGCCTTATCTGGAAGATTTAATCTTCAGTTTCAGTTGAAATATGAATGGCGACTACT	1021
Qy	694	uHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu-A	714
Db	1022	CCACGAAGACCTGGATGAAAGTGTATGACGATGATGAGAAATTCAGCCAGCTCC-CA	1080
Qy	714	snArgArgGluAspArgProIleSerPheTrrGlnLeuGlySerAsnGlnLeuGlnSerA	734
Db	1081	ACCGCGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCAGTCTA	1140
Qy	734	snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe	753
Db	1141	ACGCTGTATCTTTGGCCAGAGATGCTGCAAAACCTTGCNAAGGA-CAAGCAGAGAGGCTTC	1199
Qy	754	MetProSerIleLeuGlnAsnGluThrTrrGlyAlaLeuLeuSerGlySerProProPro	773
Db	1200	ATGCCAGACATCTTTCAGAAATGAGACTTACGAGGCCCTCTCTAGTGGCAGCCACCTCCC	1259
Qy	774	AlaGlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValGl	793
Db	1260	GCCCAGCCTGCGAGCCCGCAGCACCACCGCGCCCCCGCTTCTCCACGGAATGTGG	1319
Qy	793	yLysValGlnThrAlaSerSerAlaAsnThrLeuTrrLysThrAsnSerValSerValAs	813
Db	1320	CAAAAGTTGAGACGCTCTCTGCTAAACCTCTGTGGAAGACAAACTCTGTAAGTGTGGA	1379
Qy	813	pGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProPr	833
Db	1380	CGGTGGAAGCGCGCAGCGATCTTCTGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCC	1439
Qy	833	oLeuArgValThrSerThrAsnProLeuThrProThrProProProValAlaLysTh	853
Db	1440	TCCTTCGCTGACATCTTACCAATCCCTGACCCCGCCCGCCCGCTTCCCGTTCGCAAGAC	1499
Qy	853	rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGl	873
Db	1500	GCCCGGCTAATGAAAGCTTGGACGAGCCGAGCAAGCTTCCCGGCTGGGATCTCACA	1559
Qy	873	nIleArgProProProLeuProProGlnProProProSerArgLeuProGlnLysLysProAl	893
Db	1560	GATCAGGCCCCCACTCTGCCCCCAGCAGCGCCCGCGCTCCCGCAGAAAGAGCTGCG	1619
Qy	893	aPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA	913
Db	1620	GCCAGGGGCT-GACAAGTCCACCCCTGACTGACCAAAAGGCCAACCCGAGAGGACTGTGG	1678

QY 913 spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA 933  
 Db 1679 ATCTCTGCAACGGAAGCTCTGGGTCTCTGTCTCAATGCTATGCTGAGCCCTG 1738  
 QY 933 laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT 953  
 Db 1739 CACCATCTCTAGGAAGTCGAGGACCAAGTTGAGCCCTAAGCGGTGAAGGGCTCT 1798  
 QY 953 YrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValLleIleV 973  
 Db 1799 ATAACTGTGTGGTGACACCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCG 1858  
 QY 973 alAspGlyClnGluAspGlnGluThrTrpLleGlyHisIleAspGlyAspProGlyArgL 993  
 Db 1859 TGGACGGGGAGGAGGACGAGGAGTGTGTGATTGGCCACATTGATGAGATCTCTGTGCGCA 1918  
 QY 993 YsGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006  
 Db 1919 AAGCGCATTCCTGGGTGTCAATTGTGCACTTTATCGCTGAC 1959

## RESULT 2

ID ADB62160  
 AD B62160 standard; cDNA; 3123 BP.

AC ADB62160;

XX 04-DEC-2003 (first entry)

XX Human cDNA encoding clone CTONG20004520.

XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 146..772

FT /\*tag= a

FT /product= "Clone CTONG20004520 protein"

XX EP1308459-A2.

XX 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

XX 25-JAN-2002; 2002US-00350978.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-450961/43.

XX P-PSDB; ADB64130.

XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.

XX Claim 1; Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide

CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC is based on sequence information supplied by the European Patent Office.

XX Sequence 3123 BP; 870 A; 671 C; 610 G; 972 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4.63e-98 Length: 3123  
 Score: 110.00 Matches: 165  
 Percent Similarity: 98.80% Conservative: 0  
 Best Local Similarity: 98.80% Mismatches: 1  
 Query Match: 10.93% Indels: 2  
 DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x ADB62160 (1-3123)

QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
 Db 272 CCCCTGACCCCGACGCGCCGCCCCCTTCCCAAGACGCGCAGCGTAAATGGAAGCTTG 331  
 QY 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProLeuPro 880  
 Db 332 AGCCAGCGGAGCAAGCTGCCCCCTGGGATCTCACAGATCAGCGCCCCACCTCTGCC 391  
 QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900  
 Db 392 CCACAGCGCGCCAGCGCGCTCCCGCAGNAGAGGCTGCGCGGGGCT-GACNAGTCCAC 450  
 QY 900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920  
 Db 451 CCCACTGACCAACAAGGCGCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 510  
 QY 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGl 940  
 Db 511 GGGTCTCTGTGCCAATGCTATGGTCTCTGACAGCCCCCTGCACCCATGCTTAGGAAGTCGCA 570  
 QY 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960  
 Db 571 GGCAACCAAGTTGAAGCTTAGCGGGTGAAGCGCTCTATTAACCTGTGTGGCTGACACCC 630  
 QY 960 oAspGluLeuThrPheSerGluGlyAspValLleIleValAspGlyGluGluAspGlnGl 980  
 Db 631 CGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGGACGCGGAGGAGGACCGA 690  
 QY 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
 Db 691 GTGGTGGATTGGCCACATTGATGAGATCTCTGGTGCACAAAGCGCATTTCCCGGTGTCAAT 750  
 QY 1000 eValHisPheIleAlaAsp 1006  
 Db 751 TGTGCACTTTATCGCTGAC 769

## RESULT 3

ABK43447

ID ABK43447 standard; cDNA; 3436 BP.

XX ABK43447;

DT 05-JUN-2002 (first entry)

XX	DNA encoding novel central nervous system protein #27.	PR	12-SEP-2000; 2000US-0231968P.
DE		PR	14-SEP-2000; 2000US-0232397P.
XX		PR	14-SEP-2000; 2000US-0232398P.
KW	Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;	PR	14-SEP-2000; 2000US-0232399P.
KW	hyperproliferative disorder; neoplasia; cardiovascular disorder;	PR	14-SEP-2000; 2000US-0232400P.
KW	cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;	PR	14-SEP-2000; 2000US-0232401P.
KW	nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;	PR	14-SEP-2000; 2000US-0233063P.
KW	acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;	PR	14-SEP-2000; 2000US-0233064P.
KW	adenocarcinoma; reproductive system disorder; testicular feminisation;	PR	14-SEP-2000; 2000US-0233065P.
KW	endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;	PR	21-SEP-2000; 2000US-0234223P.
KW	respiratory disorder; renal disorder; kidney failure; blood disorder;	PR	21-SEP-2000; 2000US-0234274P.
KW	myocardial infarction; wound healing; cell proliferation; skin aging;	PR	25-SEP-2000; 2000US-0234997P.
KW	food additive; food preservative; gene therapy; gene; ss.	PR	25-SEP-2000; 2000US-0234998P.
XX		PR	26-SEP-2000; 2000US-0235484P.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834P.
XX		PR	27-SEP-2000; 2000US-0235836P.
FN	WO200155318-A2.	PR	29-SEP-2000; 2000US-0236327P.
XX		PR	29-SEP-2000; 2000US-0236367P.
PD	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236368P.
XX		PR	29-SEP-2000; 2000US-0236370P.
XX	17-JAN-2001; 2001WO-US001332.	PR	02-OCT-2000; 2000US-0236802P.
XX		PR	02-OCT-2000; 2000US-0237037P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237038P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237039P.
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000; 2000US-0237040P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239935P.
PR	16-MAR-2000; 2000US-0189874P.	PR	13-OCT-2000; 2000US-0239937P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0240960P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241221P.
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000; 2000US-0241785P.
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000; 2000US-0241786P.
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000; 2000US-0241787P.
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000; 2000US-0241808P.
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000; 2000US-0241809P.
PR	07-JUL-2000; 2000US-0216880P.	PR	01-NOV-2000; 2000US-0244617P.
PR	11-JUL-2000; 2000US-0217487P.	PR	08-NOV-2000; 2000US-0246474P.
PR	14-JUL-2000; 2000US-0217496P.	PR	08-NOV-2000; 2000US-0246475P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246476P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246477P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246478P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246523P.
PR	14-AUG-2000; 2000US-0225213P.	PR	08-NOV-2000; 2000US-0246524P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246525P.
PR	14-AUG-2000; 2000US-0225266P.	PR	08-NOV-2000; 2000US-0246526P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246527P.
PR	14-AUG-2000; 2000US-0225268P.	PR	08-NOV-2000; 2000US-0246528P.
PR	14-AUG-2000; 2000US-0225270P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0225757P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225758P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225759P.	PR	08-NOV-2000; 2000US-0246611P.
PR	18-AUG-2000; 2000US-0226279P.	PR	08-NOV-2000; 2000US-0246613P.
PR	22-AUG-2000; 2000US-0226681P.	PR	17-NOV-2000; 2000US-0249207P.
PR	22-AUG-2000; 2000US-0226688P.	PR	17-NOV-2000; 2000US-0249208P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249209P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249210P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249211P.
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000; 2000US-0249212P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249213P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249214P.
PR	01-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249215P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249216P.
PR	05-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249217P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249218P.
PR	06-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249244P.
PR	08-SEP-2000; 2000US-0231242P.	PR	17-NOV-2000; 2000US-0249245P.
PR	08-SEP-2000; 2000US-0231243P.	PR	17-NOV-2000; 2000US-0249264P.
PR	08-SEP-2000; 2000US-0231244P.	PR	17-NOV-2000; 2000US-0249265P.
PR	08-SEP-2000; 2000US-0231413P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0231414P.	PR	17-NOV-2000; 2000US-0249299P.
PR	08-SEP-2000; 2000US-0232080P.	PR	01-DEC-2000; 2000US-0249300P.
PR	08-SEP-2000; 2000US-0232081P.	PR	

PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0255967P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-581633/65.  
XX P-PSDB; AAU87117.  
XX  
XX New isolated nucleic acid encoding a protein for diagnosing, preventing,  
PT treating or ameliorating medical conditions and used as food additives or  
PT preservatives..  
XX  
XX Claim 1; SEQ ID NO 37; 837bp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC novel central nervous system protein. (I) and polypeptides (III) encoded  
CC by (I), are used to treat a medical conditions and in diagnosis of a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angioneurosis, nervous system disorders e.g. Alzheimer's disease and  
CC amyotrophic lateral sclerosis. Infections caused by bacteria, viruses  
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
CC acute kidney failure and blood related disorders e.g. myocardial  
CC infarction. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
Alignment Scores:  
Pred. No.: 5,086-98 Length: 3436  
Score: 110.00 Matches: 182  
Percent Similarity: 98.91% Conservative: 0  
Best Local Similarity: 98.91% Mismatches: 1  
Query Match: 10.93% Indels: 2  
DB: 4 Gaps: 0  
US-09-914-042-1 (1-1006) x ABK43447 (1-3436)  
Qy 612 GlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThr 631  
Db 1 GGGAACTGGATAAACAAGAGGAAAGGAGCAGCAGCCCTGCACTACTGCGCTGACC 60  
Qy 632 AspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluLeuAla 651  
Db 61 GACAATGCCGAGTGCTCAAGTTGCTCTCGGGGGAAGGCGCTTCATCGATGAGCAAC 120  
Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671  
Db 121 GAGTCAGAGAGAGACTCCCTGACATTCGCAAGCCCTCAAGCAGCAGCATGTGAGGAG 180  
Qy 672 LeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691

Db 181 CTGCTGACCCAGGCTTATCTGGAAGATTTAAATCTTCACGTTTCAGTTGAATATGAATGG 240  
Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711  
Db 241 CGACTACTCCACGAGACCTGGATGAAAGTGTATGACGACATGGATGAGAAATTCGAGCCC 300  
Qy 712 SerGlu-AsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731  
Db 301 AGTCC-CAACCGCGGGGAAAGACCGCCCATCAGCTTCTACCGAGCTGGGCTCCAACCACT 359  
Qy 731 uGlnSerAsnAlaValSerLeuAlaAatcAspAlaAlaAsnLeuAlaLysGluLysGlnA 751  
Db 360 TCAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAAACCTTGCAAGGAGAACGAG 419  
Qy 751 gAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPr 771  
Db 420 GGCATTTCATGCCAGCATCTTTCAGAGATGAGCATTCAGGAGCCCTCTCTGAGTGGAGCCC 479  
Qy 771 oProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAs 791  
Db 480 ACCTCCCGCCAGGCTGCAGCCCGCCAGCACCAGCCCGCCCGCTTCTTCCACGGAA 539  
Qy 791 nValGlyLys 794  
Db 540 TGTGGCAAA 549  
RESULT 4  
ADIS3834  
ID ADIS3834 standard; cDNA; 3436 BP.  
XX ADIS3834;  
XX  
XX 22-APR-2004 (first entry)  
XX cDNA encoding novel human protein seq id 37.  
XX  
XX neuroprotective; neurotropic; antiparkinsonian; anticonvulsant;  
KW antidiabetic; antirheumatic; antiarthritic; dermatological;  
KW antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;  
KW anti-HIV; hepatotropic; virucide; antibacterial; fungicide; respiratory;  
KW antiparasitic; muscular; gynaecological; gastrointestinal; cardiac;  
KW cardiovascular; antihypertensive; cytototoxic; antirhythmic; cardiant;  
KW nephrotropic; litholytic; cytostatic; gene therapy; neural disorder;  
KW Alzheimer's disease; Parkinson's disease; Huntington's chorea;  
KW amyotrophic lateral sclerosis; multiple sclerosis;  
KW immune system disorder; diabetes; rheumatoid arthritis;  
KW systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;  
KW inflammatory disorder; ischaemia-reperfusion injury;  
KW inflammatory bowel disease; Crohn's disease; infectious disease;  
KW HIV infection; hepatitis infection; bacterial infection;  
KW fungal infection; parasitic infection; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; pulmonary disorder;  
KW cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;  
KW renal disorder; acute glomerulonephritis; pyelonephritis;  
KW renal lithiasis; proliferative disorder; cancerous diseases; human; gene;  
XX ss.  
XX Homo sapiens.  
OS  
XX  
XX US2004018969-A1.  
XX  
XX 29-JAN-2004.  
XX  
XX 17-JAN-2001; 2001US-00764875.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215113P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216680P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

(ROSE/) ROSEN C A.  
 (RUBE/) RUBEN S M.  
 (BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI, 2004-122079/12.  
 P-PSDB; ADI54432.

New polypeptides and nucleic acid molecules, useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating medical conditions e.g. neural disorders, reproductive disorders or infectious diseases.

Claim 1; SEQ ID NO 37; 413pp; English.

XX

CC The invention describes an isolated polypeptide comprising an amino acid  
 CC sequence at least 90% identical to: a polypeptide fragment, domain,  
 CC epitope, or full-length protein of any one of 607 amino acid sequences  
 CC (i) described in the specification; a polypeptide fragment of (i), or the  
 CC encoded sequence contained in (ii), having biological activity; or a  
 CC variant, allelic variant, or a species homologue of (i). The polypeptides  
 CC and nucleic acid molecules are useful for detecting, preventing,  
 CC diagnosing, prognosticating, treating or ameliorating medical conditions  
 CC such as neural disorders, e.g. Alzheimer's disease, Parkinson's disease,  
 CC Huntington's chorea, amyotrophic lateral sclerosis or multiple sclerosis,

Alignment Scores:  
 Pred. No.: 5,08e-98 Length: 3436  
 Score: 110.00 Matches: 182  
 Percent Similarity: 98.91% Conservative: 0  
 Best Local Similarity: 98.91% Mismatches: 1  
 Query Match: 10.93% Indels: 2  
 DB: 12 Gaps: 0

US-09-914-042-1 (1-1006) x ADI53834 (1-3436)

Qy 612 GlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThr 631  
 Db 1 GGGAACTGGATAAACAGACAGGGAAGCGACAGCCCTGCACCTACTGCTGCTGACC 60  
 Qy 632 AspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsn 651  
 Db 61. GACAATGCCAGTGCTCAAGTTGCTCTCGGGGGAAGGCTCCATCGATAGCAAAAC 120  
 Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671  
 Db 121 GAGTCAGAGAGACTCCGCTGGACATTCGCAAGCGCTCAAGACGACGACTGTGAGGAG 180  
 Qy 672 LeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691  
 Db 181 CTGCTGACCCAAAGCTTATCTGGAGATTATTTCTACGTTTACGTTGAATATGAATGG 240  
 Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711  
 Db 241 CGACTACTCCACGAAGACCTGGATGAAAGTATGATGACGATGATGATGATGATGATGATG 300  
 Qy 712 SerGlu-AsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731  
 Db 301 AGTCC-CAACCGCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAACCGACT 359  
 Qy 731 uGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnAr 751  
 Db 360 TCAGTCTAACCTGTATCTTTGGCAGAGATGCTGCAAACTTGCACAGGAGAGACGAG 419  
 Qy 751 gAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuSerGlySerPr 771  
 Db 420 GGCTTTTCATGCCAGCATCTTGCAAGATGAGACTTACGAGGCCCTCTCTGAGTGGCAGGCC 479  
 Qy 771 oProProGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAs 791  
 Db 480 ACCTCCCGCCGAGCTGAGCGCCCGACGACCAACGAGCGCCCGGCTTCTCTCCACGGA 539  
 Qy 791 nValGlyLys 794  
 Db 540 TGTGGCANA 549

RESULT 5  
 ABK43768  
 ID ABK43768 standard; cDNA; 3413 BP.

XX AC ABK43768;

XX DT 05-JUN-2002 (first entry)

XX DE DNA encoding novel central nervous system protein #348.

XX KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; Gene therapy; gene; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200155318-A2.  
 XX 02-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US001332.  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-019076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 14-AUG-2000; 2000US-0221829P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225288P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.





811	-GGTCCTGCAGCCCTCGACCCCATCGCTAGGAAGTCGCGAGGCAACCAAGTTGAAGCCTA	869	PR	14-AUG-2000;	2000US-02252131P.
			PR	14-AUG-2000;	2000US-02252144P.
947	ysArgVallysAlaLeuTyraenCysValAlaAspAsnProAspGluLeuThrPheSerG	967	PR	14-AUG-2000;	2000US-0225266P.
870	AGCGGTGAAGCGCTCTATAACTGTGTGGCTGACAAACCCGATGCTCACCTTCTCCG	929	PR	14-AUG-2000;	2000US-0225267P.
			PR	14-AUG-2000;	2000US-0225268P.
967	luGlyAspValIlelleValAspGlyGluGluAspGlnGluTrpTrpIleGlyHisIleA	987	PR	14-AUG-2000;	2000US-0225270P.
			PR	14-AUG-2000;	2000US-0225447P.
930	AGGGGATGTGATCATCTGTGACGGGAGGAGGACGAGGTGGTGGATGGCCACATG	989	PR	14-AUG-2000;	2000US-0225757P.
			PR	14-AUG-2000;	2000US-0225758P.
			PR	14-AUG-2000;	2000US-0225759P.
987	spGlyAspProGlyVargLysGlyAlaPheProValSerPheValHisPheIleAlaAsp	1006	PR	18-AUG-2000;	2000US-0226279P.
			PR	22-AUG-2000;	2000US-0226681P.
990	ATGGAGATCTGTGTGCAAGGCGCATTCCTCCGGTGTCACTTTATCGCTGAC	1048	PR	22-AUG-2000;	2000US-0226686P.
			PR	22-AUG-2000;	2000US-0227182P.
			PR	23-AUG-2000;	2000US-0227009P.
			PR	30-AUG-2000;	2000US-0228924P.
			PR	01-SEP-2000;	2000US-0229287P.
			PR	01-SEP-2000;	2000US-0229343P.
			PR	01-SEP-2000;	2000US-0229344P.
			PR	01-SEP-2000;	2000US-0229345P.
			PR	05-SEP-2000;	2000US-0229509P.
			PR	05-SEP-2000;	2000US-0229513P.
			PR	06-SEP-2000;	2000US-0230437P.
			PR	06-SEP-2000;	2000US-0230438P.
			PR	08-SEP-2000;	2000US-0231242P.
			PR	08-SEP-2000;	2000US-0231243P.
			PR	08-SEP-2000;	2000US-0231244P.
			PR	08-SEP-2000;	2000US-0231413P.
			PR	08-SEP-2000;	2000US-0231414P.
			PR	08-SEP-2000;	2000US-0232080P.
			PR	08-SEP-2000;	2000US-0232081P.
			PR	12-SEP-2000;	2000US-0231968P.
			PR	14-SEP-2000;	2000US-0232397P.
			PR	14-SEP-2000;	2000US-0232398P.
			PR	14-SEP-2000;	2000US-0232399P.
			PR	14-SEP-2000;	2000US-0232400P.
			PR	14-SEP-2000;	2000US-0232401P.
			PR	14-SEP-2000;	2000US-0233063P.
			PR	14-SEP-2000;	2000US-0233064P.
			PR	14-SEP-2000;	2000US-0233065P.
			PR	21-SEP-2000;	2000US-0234223P.
			PR	21-SEP-2000;	2000US-0234274P.
			PR	25-SEP-2000;	2000US-0234397P.
			PR	25-SEP-2000;	2000US-0234998P.
			PR	26-SEP-2000;	2000US-0235484P.
			PR	27-SEP-2000;	2000US-0235834P.
			PR	27-SEP-2000;	2000US-0235836P.
			PR	29-SEP-2000;	2000US-0236327P.
			PR	29-SEP-2000;	2000US-0236367P.
			PR	29-SEP-2000;	2000US-0236368P.
			PR	29-SEP-2000;	2000US-0236369P.
			PR	29-SEP-2000;	2000US-0236370P.
			PR	02-OCT-2000;	2000US-0236802P.
			PR	02-OCT-2000;	2000US-0237037P.
			PR	02-OCT-2000;	2000US-0237038P.
			PR	02-OCT-2000;	2000US-0237039P.
			PR	02-OCT-2000;	2000US-0237040P.
			PR	13-OCT-2000;	2000US-0239935P.
			PR	13-OCT-2000;	2000US-0239937P.
			PR	20-OCT-2000;	2000US-0240960P.
			PR	20-OCT-2000;	2000US-0241221P.
			PR	20-OCT-2000;	2000US-0241785P.
			PR	20-OCT-2000;	2000US-0241786P.
			PR	20-OCT-2000;	2000US-0241787P.
			PR	20-OCT-2000;	2000US-0241808P.
			PR	20-OCT-2000;	2000US-0241809P.
			PR	20-OCT-2000;	2000US-0241826P.
			PR	01-NOV-2000;	2000US-0244617P.
			PR	08-NOV-2000;	2000US-0246474P.
			PR	08-NOV-2000;	2000US-0246475P.
			PR	08-NOV-2000;	2000US-0246476P.
			PR	08-NOV-2000;	2000US-0246477P.
			PR	08-NOV-2000;	2000US-0246478P.



PT New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.

PS Claim 1; SEQ ID NO 27010; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 442 BP; 137 A; 97 C; 101 G; 106 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 3.5e-67 Length: 442  
 Score: 78.00 Matches: 78  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.75% Indels: 0  
 DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x ACH39798 (1-442)

Qy 365 CysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGlu 384  
 Db 38 TGCCTTTGACCTCATTCATGACAGAACTTACCACCTTTCAAGCTGAAGATGACAGGAA 97  
 Qy 385 CysGlnIleTrpMetSerValLeuGlnAsnSerIysGluGluAlaLeuAsnAsnAlaPhe 404  
 Db 98 TGTCAATATGATGATCTGTCTGCTGCACAAATAGCAAGAAGAGCTTTAAACAATGCATTT 157  
 Qy 405 LysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIle 424  
 Db 158 AAGGGGGATGACATATCTGGAGAAATACATCTGTCAGAACTGACAAAGAGATCATC 217  
 Qy 425 SerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaPro 442  
 Db 218 TCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGTGACTGTGGGGCGCCA 271

RESULT 8

AAV59105  
 ID AAV59105 standard; cDNA; 5954 BP.

XX AAV59105;

XX 17-OCT-2003 (revised)

DT 20-NOV-1998 (first entry)

XX Zebrafish differentiation enhancing factor 2 gene.

XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
 KW C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
 KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
 KW nervous system.

OS Danio rerio.

XX Key Location/Qualifiers  
 FH 433. .3381  
 FT CDS

FT /\*tag= a  
 FT /product= "DEF2 protein"

XX WO9836065-A1.

XX 20-AUG-1998.

XX 13-FEB-1998; 98WO-US002724.

XX 14-FEB-1997; 97US-0038191P.

XX (DAND ) DANA FARBER CANCER INST INC.

XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;

XX WPI; 1998-467173/40.

XX P-PSDB; AAW77288.

XX New nucleic acid encoding differentiation enhancing factor - used  
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
 PT tumours and neurological disease.

XX Claim 7; Fig 14; 203pp; English.

XX The differentiation enhancing factors (DEF), comprise at least one each  
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
 CC are mediators of SH3-domain dependent signalling and may be involved in  
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
 CC endocytosis or adhesion, migration, proliferation and differentiation of  
 CC cells. Typical applications of DEF and agents that modulate interaction  
 CC between the protein and it's ligand, or of nucleic acid expressing them,  
 CC are treatment of hyperplastic and neoplastic disease (a wide range of  
 CC solid tumours and leukaemias), including metastases; for in vitro  
 CC induction of differentiation of neural crest cells to neurons, glial  
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
 CC the nervous system (e.g. treatment of traumatic injury, stroke,  
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
 CC standardise OS field)

XX Sequence 5954 BP; 1731 A; 1382 C; 1372 G; 1469 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.32e-56 Length: 5954  
 Score: 68.00 Matches: 68  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.76% Indels: 0  
 DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AAV59105 (1-5954)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGlu 190  
 Db 943 CACGGAATGATCCGACGAGATCAGCGGACAGATAGCAGAAGAGATGAAAAAG 1002  
 Qy 191 ArgArgPheGlnLeuGlnMetCysGluTrpLeuLeuValAsnGluIleLysIle 210  
 Db 1003 CGCGGTTTCCTCCAGCTTCAGATGTGTGAGTACCTCTCAAGTCAATGAATCAAGATC 1062  
 Qy 211 LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTrpPheHisAlaGlnCysAsnPhe 230  
 Db 1063 AAAAAAGGTGTCGACCTGCTCCAGAAATCTCATCAATATCTCCACGCACAGTCACTTC 1122  
 Qy 231 PheGlnAspGlyLeuLysAlaVal 238  
 Db 1123 TTTCAGGATGCTCTCAAAAGCGGTG 1146

RESULT 9

AAV58478

ID AAV58478 standard; cDNA; 563 BP.

XX

AC AAS58478;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE cDNA #1154 encoding portion of a human colon tumour protein.  
 XX  
 KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200173027-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 XX 22-MAR-2001; 2001WO-US009246.  
 XX  
 XX 24-MAR-2000; 2000US-0191597P.  
 PR 04-MAY-2000; 2000US-0202024P.  
 PR 05-MAY-2000; 2000US-0202189P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Meagher MJ, Xu J, King GE;  
 XX  
 DR WPI; 2001-611627/70.  
 XX  
 XX New colon tumor proteins and related nucleic acid, useful for treatment,  
 PT prevention, diagnosis and monitoring of cancer.  
 XX  
 PS Claim 4; Page 238; 299pp; English.  
 XX  
 CC Th present invention relates to the isolation of novel cDNA sequences  
 CC encoding for at least an immunogenic portion of human colon tumour  
 CC proteins. The sequences of the invention are useful in pharmaceutical  
 CC compositions and vaccines for the prevention and treatment of cancers  
 CC such as colon cancer. They are also useful for the diagnosis and  
 CC monitoring of such cancers. Antibodies to the colon tumour proteins and  
 CC antigen presenting cells that express polynucleotides encoding colon  
 CC tumour proteins can be used to inhibit the development of cancers. T-  
 CC cells that react specifically with colon tumour proteins are useful for  
 CC removing tumour cells from samples (e.g. blood) and for cancer treatment.  
 CC The polynucleotides sequences are also useful in gene therapy. AAS57325-  
 CC AAS58880 represent the cDNA sequences of the invention that encode for  
 CC portions of human colon tumour proteins  
 XX  
 SQ Sequence 563 BP; 163 A; 131 C; 140 G; 119 T; 0 U; 10 Other;  
  
 Alignment Scores:  
 Pred. No.: 2.24e-37 Length: 563  
 Score: 48.00 Matches: 131  
 Percent Similarity: 96.32% Conservative: 0  
 Best Local Similarity: 96.32% Mismatches: 2  
 Query Match: 4.77% Indels: 5  
 DB: 4 Gaps: 0  
  
 US-09-914-042-1 (1-1006) x AAS58478 (1-563)  
  
 QY 401 AenAenAlaPheLysGlyAspAspAenThrGlyGluAenAenIleValGlnGluLeuThr 420  
 DB 2 AACATGCAATTTAAAGGGGATGACAATACTGGAGAAATAAATCATCGTCCAAAGACTGACA 61  
  
 QY 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAenAepValCysCysAspCysG1 440  
 DB 62 AAGGAGATCACTCAGAGTGTCT-GAGGATGACGGGCAATGACGTCTGTGACTGTGG 120  
  
 QY 440 YAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460  
 DB 121 GGGCGCAGATCTCATATGCTTTCCACCAACCTGGGCATCTCGACCTGCATCGAGTGTTC 180  
  
 QY 460 rGlyIleHisArgGluLeuGlyValHisTrpSerPro-MetGlnSerLeuThrLeuAspV 480  
 DB 181 CGGAATCCACCGAGAGCTGGGGGTTTCATTATTCCAG-GATGCGATCCCTGACCTTAGATG 239

QY 480 alLeuGlyThrSerGluLeuLeuAlaLysAenIleGlyAsnAlaGlyPheAsnGluI 500  
 DB 240 TACTGGGACATCTGAGCTGCTGCCAAGATATTGGGAATGCAGGCTTTAATGAGA 299  
  
 QY 500 leMetGluCysCysLeuProAlaGluAepSerValLysProAsnProGlySer-AspMet 519  
 DB 300 TCATGGAATGTGCTACCACTGAGGACTCAGTCAAAACCCCAACCCAGGACGACATG 359  
  
 QY 520 AsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533  
 DB 360 AATGCAAGAAAGGACTACATCACAGCCAAAGTACATCGAGAGG 401  
  
 RESULT 10  
 AAV59106  
 ID AAV59106 standard; cDNA; 4595 BP.  
 XX  
 AC AAV59106;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 20-NOV-1998 (first entry)  
 XX  
 DE Zebrafish differentiation enhancing factor 3 gene.  
 XX  
 KW ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
 KW C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
 KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
 KW nervous system.  
 XX  
 OS Danio rerio.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 300..3011  
 FT /\*tag= a  
 FT /product= "DEF3 protein"  
 XX  
 FN WO9836065-A1.  
 XX  
 PD 20-AUG-1998.  
 XX  
 PP 13-FEB-1998; 98WO-US002724.  
 XX  
 PR 14-FEB-1997; 97US-0038191P.  
 XX  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 PI Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;  
 XX  
 DR WPI; 1998-467173/40.  
 DR P-PSDB; AAW59106.  
 XX  
 PT New nucleic acid encoding differentiation enhancing factor - used  
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
 PT tumours and neurological disease.  
 XX  
 PS Claim 7; Fig 15; 203pp; English.  
 XX  
 CC The differentiation enhancing factors (DEF), comprise at least one each  
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
 CC are mediators of SH3-domain dependent signalling and may be involved in  
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
 CC endocytosis or adhesion, migration, proliferation and differentiation of  
 CC cells. Typical applications of DEF and agents that modulate interaction  
 CC between the protein and it's ligand, or of nucleic acid expressing them,  
 CC are treatment of hyperplastic and neoplastic disease (a wide range of  
 CC solid tumours and leukemias), including metastases; for in vitro  
 CC induction of differentiation of neural crest cells to neurons, glial  
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
 CC the nervous system (e.g. treatment of traumatic injury, stroke,  
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
 CC standardise OS field)  
 XX

SQ Sequence 4595 BP; 1378 A; 1014 C; 1030 G; 1173 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.15e-20 Length: 4595  
Score: 32.00 Matches: 32  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.18% Indels: 0  
DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AAV59106 (1-4595)

QY 125 PheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLys 144

DB 672 TTTCACATAGACAGCTCTCAAGGGAGACCTCAAGGAGTCAAGGGGATTTGAAAAG 731

QY 145 ProPheAspLysAlaTrpLysAspTyrGluThrLys 156

DB 732 CCATTTGTAAGCATGGAAGATTATGAAACAAA 767

## RESULT 11

ABA08628/c

ID ABA08628 standard; cDNA; 409 BP.

AC ABA08628;

XX 11-JAN-2002 (first entry)

XX Human GTPase activating protein homologue-encoding cDNA, SEQ ID NO:404.

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;  
KW antifungal; vulnery; antiulcer; ss.

XX Homo sapiens.

OS WO200157188-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003800.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX P-PSDB; ABB11384.

XX Human proteins and DNA encoding sequences useful for preventing, treating  
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
PT and cancer.

PS Claim 1; Page 485; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides

CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis; cancer cell proliferation or metastasis.

CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g. asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.

CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness.

CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention

XX SQ Sequence 409 BP; 98 A; 107 C; 74 G; 130 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.1e-19 Length: 409  
Score: 30.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.98% Indels: 0  
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x ABA08628 (1-409)

QY 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLys 146

DB 133 TTGGATTCCTTTGTTAAAGGAGACCTAAAGGAGTCAAGGAGATCTCAAGAGGCATTT 74

QY 147 AspLysAlaTrpLysAspTyrGluThrLys 156

DB 73 GACAAAGCCTGGAAGACTATGAGACAAG 44

RESULT 12

ADD71214

ID ADD71214 standard; cDNA; 3775 BP.

XX AC ADD71214;

XX 15-JAN-2004 (first entry)

XX Human intracellular signalling molecule INTS1G-51 cDNA SEQ ID NO:103.  
KW human; intracellular signalling molecule; INTS1G; cytostatic;  
KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;  
KW neuroprotective; cerebroprotective; anti-HIV; antiasthmatic;  
KW antiinflammatory; thymimetic; gene therapy;  
KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;  
KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;  
KW immune disorder; inflammatory disorder; AIDS; allergy;  
KW developmental disorder; Hypothyroidism; Cushing's syndrome;  
KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.

OS Homo sapiens.  
 PN WO2003039348-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 PF 07-NOV-2002; 2002WO-US036151.  
 XX  
 PP 09-NOV-2001; 2001US-0344472P.  
 PR 30-NOV-2001; 2001US-0334558P.  
 PR 14-DEC-2001; 2001US-0340296P.  
 PR 21-DEC-2001; 2001US-0343557P.  
 PR 18-JAN-2002; 2002US-0350420P.  
 PR 25-JAN-2002; 2002US-0351927P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Kable AE, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;  
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;  
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;  
 PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;  
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;  
 PI Lee SY, Tran UK, Elliott VS, Sprague WM, Tang YT, Zebajadian Y;  
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;  
 PI Zheng W;  
 XX  
 DR WPI; 2003-441441/41.  
 DR P-PSDB; ADD71162.  
 XX  
 XX New human intracellular signalling molecules (INTSIG)), useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or  
 PT infections.  
 XX  
 PS Claim 5; SEQ ID NO 103; 363pp; English.  
 XX  
 CC The present invention describes human intracellular signalling molecules  
 CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins  
 CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,  
 CC antiinflammatory and thymomimetic activities, and can be used in the  
 CC diagnosis, treatment and prevention of diseases or conditions associated  
 CC with the decreased expression or overexpression of INTSIG, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),  
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),  
 CC immune/inflammatory (e.g. AIDS, allergies), reproductive and vesicle-trafficking  
 CC disorders, or infections. They are also useful in assessing the effects  
 CC of exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of INTSIG. The INTSIG or its fragments are useful in screening  
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
 CC or in altering the expression of the target polynucleotide and compounds  
 CC that specifically bind to or modulate the activity of the polypeptide.  
 CC The present sequence encodes human INTSIG-51 from the present invention.  
 XX  
 SQ Sequence 3775 BP; 1123 A; 989 C; 873 G; 790 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 9,13e-19 Length: 3775  
 Score: 30.00 Matches: 30  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.98% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-914-042-1 (1-1006) x ADD71214 (1-3775)  
 QY 127 LeuApsSerLeuLeuLeuGlyAspLeuLeuGlyValLeuGlyAspLeuLeuLysProphe 146  
 DB 696 TTGGATTCCTTTGTTAAAGGAGACCTTAAGGGAGTCAAAGGAGATCTCAAGAAGCCATTT 755  
 QY 147 AspLysAlaTrpLysAspTyrGluThrLys 156

Db 756 GACAAGCCTGGAAGATTATGACACAAG 785  
 RESULT 13  
 ADD71207  
 ID ADD71207 standard; cDNA; 3946 BP.  
 XX  
 AC ADD71207;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Human intracellular signalling molecule INTSIG-44 cDNA SEQ ID NO:96.  
 XX  
 KW human: intracellular signalling molecule; INTSIG: cytostatic;  
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; neurotropic;  
 KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;  
 KW antiinflammatory; thymomimetic; gene therapy;  
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;  
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;  
 KW immune disorder; inflammatory disorder; AIDS; allergy;  
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;  
 KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003039348-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 XX 07-NOV-2002; 2002WO-US036151.  
 XX  
 PF 09-NOV-2001; 2001US-0344472P.  
 PR 30-NOV-2001; 2001US-0334558P.  
 PR 14-DEC-2001; 2001US-0340296P.  
 PR 21-DEC-2001; 2001US-0343557P.  
 PR 18-JAN-2002; 2002US-0350420P.  
 PR 25-JAN-2002; 2002US-0351927P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Kable AE, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;  
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;  
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;  
 PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;  
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;  
 PI Lee SY, Tran UK, Elliott VS, Sprague WM, Tang YT, Zebajadian Y;  
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;  
 PI Zheng W;  
 XX  
 DR WPI; 2003-441441/41.  
 DR P-PSDB; ADD71155.  
 XX  
 XX New human intracellular signalling molecules (INTSIG)), useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or  
 PT infections.  
 XX  
 PS Claim 5; SEQ ID NO 96; 363pp; English.  
 XX  
 CC The present invention describes human intracellular signalling molecules  
 CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins  
 CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,  
 CC antiinflammatory and thymomimetic activities, and can be used in the  
 CC diagnosis, treatment and prevention of diseases or conditions associated  
 CC with the decreased expression or overexpression of INTSIG, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),  
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),  
 CC immune/inflammatory (e.g. AIDS, allergies), reproductive and vesicle-trafficking  
 CC disorders, or infections. They are also useful in assessing the effects  
 CC of exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of INTSIG. The INTSIG or its fragments are useful in screening  
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
 CC or in altering the expression of the target polynucleotide and compounds  
 CC that specifically bind to or modulate the activity of the polypeptide.  
 CC The present sequence encodes human INTSIG-51 from the present invention.  
 XX  
 SQ Sequence 3775 BP; 1123 A; 989 C; 873 G; 790 T; 0 U; 0 Other;

CC sequences of INTSIG. The INTSIG or its fragments are useful in screening  
CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
CC or in altering the expression of the target polynucleotide and compounds  
CC that specifically bind to or modulate the activity of the polypeptide.  
CC The present sequence encodes human INTSIG-44 from the present invention.  
XX  
SQ Sequence 3946 BP; 1165 A; 1064 C; 902 G; 815 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	9,52e-19	Length:	3946
Score:	30.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.98%	Indels:	0
DB:	10	Gaps:	0

US-09-914-042-1 (1-1006) x ADD71207 (1-3946)

QY 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe 146  
|||||  
Db 696 TTGGATTCTTTGTTAAAGGAGACCTAAAGGAGTCAAGGAGATCTCAAGAGCCATT 755  
|||||  
QY 147 AspLysAlaTdpLysAspTyrGluThrLys 156  
|||||  
Db 756 GACAAAGCTGGAAGATTTAGACAAAG 785  
|||||

## RESULT 14

ADJ51008  
ID ADJ51008 standard; cDNA; 3974 BP.

AC ADJ51008;

XX 06-MAY-2004 (first entry)

XX Human cDNA encoding NOV31a.

XX Human; ss; gene; NOVX; autoimmune disease; Alzheimer's disease; stroke;  
KW allergy; Parkinson's disease; Huntington's disease; multiple sclerosis;  
KW anxiety; pain; diabetes; graft versus host disease; pancreatitis;  
KW obesity; ulcer; anaemia; cancer; viral infection; bacterial infection;  
KW parasitic infection.

XX Homo sapiens.

XX US2004030096-A1.

XX 12-FEB-2004.

XX 01-AUG-2002; 2002US-00210281.

XX 02-AUG-2001; 2001US-0309501P.

XX 03-AUG-2001; 2001US-0310291P.

XX 08-AUG-2001; 2001US-0310951P.

XX 09-AUG-2001; 2001US-0311292P.

XX 13-AUG-2001; 2001US-0311979P.

XX 14-AUG-2001; 2001US-0312203P.

XX 17-AUG-2001; 2001US-0313201P.

XX 20-AUG-2001; 2001US-0313643P.

XX 20-AUG-2001; 2001US-0313702P.

XX 21-AUG-2001; 2001US-0314031P.

XX 23-AUG-2001; 2001US-0314466P.

XX 28-AUG-2001; 2001US-0315403P.

XX 29-AUG-2001; 2001US-0315853P.

XX 05-MAR-2002; 2002US-0361775P.

XX 05-MAR-2002; 2002US-0361832P.

XX (GORM/) GORMAN L.

PA (ZERH/) ZERHUSEN B D.

PA (EDIN/) EDINGER S R.

PA (PADI/) PADIGARU M.

PA (GUOX/) GUO X.

PA (KEKU/) KEKUDA R.

PA (ZHON/) ZHONG M.

PA (PATT/) PATTURAJAN M.  
PA (MILL/) MILLER C E.  
PA (JIWW/) JI W.  
PA (PENA/) PENA C E A.  
PA (BURG/) BURGESS C E.  
PA (SCIO/) SCIORE P.  
PA (STON/) STONE D J.  
PA (TAUP/) TAUPIER R J.  
PA (CASM/) CASMAN S J.  
PA (ROTH/) ROTHENBERG M E.  
PA (MALY/) MALYANKAR U M.  
PA (BOLD/) BOLDOG F L.

Gorman L, Zerrhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R;  
Zhong M, Patturajan M, Miller CE, Ji W, Pena CE, Burgess CE;  
Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME;  
Malyankar UM, Boldog FL;

WPI; 2004-168942/16.  
P-PSDB; ADJ51009.

New NOVX polypeptides and polynucleotides, useful in diagnosing, treating  
or preventing diseases or conditions, e.g. autoimmune disease, treating  
Alzheimer's disease, diabetes, graft versus host disease, cancer or viral  
or bacterial infections.

Claim 20; SEQ ID NO 73; 342pp; English.

The invention relates to an isolated NOVX polypeptide (of 44 disclosed)  
comprising its mature form, a sequence having at least 98% sequence  
identity to NOVX or a sequence comprising one or more conservative  
substitutions in the amino acid sequence of NOVX. Also included are a  
composition comprising NOVX and a carrier, a kit comprising, in one or  
more containers, the composition, a method of identifying an agent that  
binds to NOVX, a method for identifying a potential therapeutic agent for  
use in treatment of a pathology related to aberrant expression or  
aberrant physiological interactions of NOVX, a method for screening for a  
modulator of activity of or of latency or predisposition to a pathology  
associated with NOVX, a method for modulating the activity of NOVX, a  
method of treating or preventing a pathology associated with NOVX or a  
pathological state in a mammal, an isolated nucleic acid molecule  
encoding a NOVX protein, a vector comprising the nucleic acid molecule,  
a cell comprising the vector, an antibody that immunospecifically binds  
to NOVX, a method for determining the presence or amount of NOVX or the  
nucleic acid molecule in a sample, a method for determining the presence  
of or predisposition to a disease associated with altered levels of  
expression of NOVX or the nucleic acid molecule in a first mammalian  
subject and a method of producing NOVX (comprising culturing the cell  
under conditions that lead to expression of the polypeptide). NOVX is  
useful in the manufacture of a medicament for treating a syndrome  
associated with a human disease associated with NOVX. The polypeptides  
and nucleic acid molecules are useful in diagnosing, treating or  
preventing diseases or conditions, e.g. autoimmune disease, Alzheimer's  
disease, stroke, allergies, Parkinson's disease, Huntington's disease,  
multiple sclerosis, anxiety, pain, diabetes, graft versus host disease,  
pancreatitis, obesity, ulcers, anaemia, cancer, viral or bacterial and  
parasitic infections (many more diseases and disorders are listed in the  
specification). The present sequence encodes a NOVX protein.

SQ Sequence 3974 BP; 1208 A; 1019 C; 888 G; 859 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	9,59e-19	Length:	3974
Score:	30.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.98%	Indels:	0
DB:	12	Gaps:	0

US-09-914-042-1 (1-1006) x ADJ51008 (1-3974)

QY 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe 146  
|||||

Db 699 TTGGATTCTTTGTTAAAGGAGACCTAAAGGAGTCTAAAGGAGATCTCAAGAGCCATT 758  
Qy 147 AsplysAlaTrpLysAspTyrGluThrLys 156  
Db 759 GACAAAGCCTGGAAGATATTGAGACAAA 788

RESULT 15  
AAV59104  
ID AAV59104 standard; cDNA; 4382 BP.  
XX AAV59104;  
XX AC  
XX XX  
DT 17-OCT-2003 (revised)  
DT 20-NOV-1998 (first entry)  
XX  
DE Zebrafish differentiation enhancing factor 1 gene.  
XX  
KW ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
KW C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
KW nervous system.  
XX  
OS Danio rerio.  
XX  
XX Key Location/Qualifiers  
FH 351..3806  
FT CDS /\*tag= a  
FT /product= "DEF1 protein"  
XX  
XX WO9836065-A1.  
XX  
XX 20-AUG-1998.  
XX  
PF 13-FEB-1998; 98WO-US002724.  
XX  
PR 14-FEB-1997; 97US-0038191P.  
XX  
XX (DAND ) DANA FARBER CANCER INST INC.  
XX  
XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;  
XX  
XX WPI; 1998-467173/40.  
DR P-PSDB; AAW77287.  
XX  
PT New nucleic acid encoding differentiation enhancing factor - used  
PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
PT tumours and neurological disease.  
XX  
XX Claim 7; Fig 13; 203pp; English.  
XX  
CC The differentiation enhancing factors (DEF), comprise at least one each  
CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
CC are mediators of SH3-domain dependent signalling and may be involved in  
CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
CC endocytosis or adhesion, migration, proliferation and differentiation of  
CC cells. Typical applications of DEF and agents that modulate interaction  
CC between the protein and it's ligand, or of nucleic acid expressing them,  
CC are treatment of hyperplastic and neoplastic disease (a wide range of  
CC solid tumours and leukaemias), including metastases; for in vitro  
CC induction of differentiation of neural crest cells to neurons, glial  
CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
CC the nervous system (e.g. treatment of traumatic injury, stroke, in  
CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 4382 BP; 1222 A; 1132 C; 1098 G; 930 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1.05e-18 Length: 4382  
Score: 30.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.98% Indels: 0  
DB: 2 Gaps: 0  
US-09-914-042-1 (1-1006) x AAV59104 (1-4382)  
Qy 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe 146  
Db 789 CTGGACTCTCTGCTCAAGGAGATCTAAAGGAGTGAAGGGGACCTTAAAAAGCCTTTC 848  
Qy 147 AsplysAlaTrpLysAspTyrGluThrLys 156  
Db 849 GACAAAGCCTGGAAGACTATGAAACCAAG 878

Search completed: August 4, 2005, 19:17:20  
Job time : 1282 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 14:05:47 ; Search time 10576 Seconds

(without alignments)  
4609.110 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGPGRKGAPVPSVHFHTAD 1006

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9408497

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool\_h/US09914042/Runat\_04082005\_090746\_2657/app\_query.fasta\_1.1159  
-DB=GenEmbl -QWMT=fastcap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042 @CGN 1 1 6279 @runat\_04082005\_090746\_2657 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=110 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	717	71.3	5711	9	AB007860 Homo sapi
2	525	52.2	5544	9	BC063308 Homo sapi
3	297	29.5	4486	6	I74317 Sequence 15
4	257	25.5	4328	6	I74314 Sequence 12

5	161	16.0	4713	10	AK172944	AK172944 Mus muscu
6	161	16.0	5072	10	BC080847	BC080847 Mus muscu
7	130	12.9	2109	6	CQ720712	CQ720712 Sequence
8	110	10.9	3123	6	AX746789	AX746789 Sequence
9	110	10.9	3123	9	AK091121	AK091121 Homo sapi
10	110	10.9	3870	9	AK124900	AK124900 Homo sapi
11	108	10.7	858	5	CR385625	CR385625 Gallus ga
12	81	8.1	171653	9	AC080162	AC080162 Homo sapi
13	71	7.1	2157	9	AK123185	AK123185 Homo sapi
14	68	6.8	2949	6	AR243296	AR243296 Sequence
15	68	6.8	5954	6	AR243295	AR243295 Sequence
16	55	5.5	158388	2	AC102852	AC102852 Mus muscu
17	55	5.5	241137	2	AC112078	AC112078 Rattus no
18	55	5.5	269229	2	AC097551	AC097551 Rattus no
19	49	4.9	85411	9	AC097782	AC097782 Homo sapi
20	49	4.9	198935	2	AC102817	AC102817 Mus muscu
21	49	4.9	241137	2	AC112078	AC112078 Rattus no
22	48	4.8	563	6	AX261503	AX261503 Sequence
23	46	4.6	3989	5	BC070750	BC070750 Xenopus l
24	42	4.2	191827	9	AC093904	AC093904 Homo sapi
25	37	3.7	1721	5	CR387820	CR387820 Gallus ga
26	34	3.4	1423	10	BC028949	BC028949 Mus muscu
27	34	3.4	65282	2	AC116865	AC116865 Mus muscu
28	34	3.4	177850	2	AC113090	AC113090 Mus muscu
29	32	3.2	2712	6	AR243298	AR243298 Sequence
30	32	3.2	4595	6	AR243297	AR243297 Sequence
31	30	3.0	706	5	BX950804	BX950804 Gallus ga
32	30	3.0	2583	6	CQ716704	CQ716704 Sequence
33	30	3.0	3456	6	AR243294	AR243294 Sequence
34	30	3.0	4148	10	AK122477	AK122477 Mus muscu
35	30	3.0	4150	10	AF075462	AF075462 Mus muscu
36	30	3.0	4370	10	AF075461	AF075461 Mus muscu
37	30	3.0	4382	6	AR243293	AR243293 Sequence
38	30	3.0	4434	10	BC048818	BC048818 Mus muscu
39	30	3.0	5330	4	AF112886	AF112886 Bos tauru
40	30	3.0	5330	6	AR243292	AR243292 Sequence
41	30	3.0	209698	5	CR381585	CR381585 Zebrafish
42	27	2.7	162648	5	AL935033	AL935033 Zebrafish
43	27	2.7	212604	5	AL805945	AL805945 Zebrafish
44	24	2.4	884	5	CF523567	CF523567 Gallus ga
45	24	2.4	171767	10	AC134601	AC134601 Mus muscu

#### ALIGNMENTS

RESULT 1	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	linear	PRI 10-JAN-2004
AB007860	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
LOCUS	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
DEFINITION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
ACCESSION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
VERSION	AB007860.1	GI:2662080			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					

## source

```

1. 5711
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG1091"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. 5711
/gene="KIAA0400"
<293...3361
/gene="KIAA0400"
/note="Start codon is not identified"
/codon_start=1
/protein_id="BAA23696.2"
/db_xref="GI:40788243"
/translation="GGGAPAAVRQPRRAEAMPDQTSVSEFVAETHEDYKAPTASSFTT
RTACRNTVAIAIEALVDVRLVYKMKSVKAINSSGLAHVENEBOYTOALEKFGNGC
VCRDPDLSGSAFLKFSVFTKELTALFKNLIOMNNIISFPLDSLKGLDKGVKGLKK
PFDAKWDYETIKPIEKEKEHAKLGMIRTEISGAIEAEMEKERRFFLOQCEYL
LKVNEIKIKGVLDLQNLIKYPHACQNFQDGLKAVESLKPSIETLSDTLHIKOQAD
EERQLQRLDILKSAQVEQKDSQIRQSTAYSLHQPQGNKEHGTENRGLYKSDG
IRKWKQRKCSVKQFLTISHGTANRPAPKLNLLTCQVKNPEBKCFDLISHDRTH
FOADEBOCQIWMVLQNSKEBALNNAFKGDNDGNNI VOELTKETI SEVORTGND
VCCDGA PDPTWLTNLGILTCIECSGIHRELVGHVSRMQSLTLDVLGTSLLAKNI
GNAEFMECLPAEDSVKPNPGSDMNARKDYITAKYIERRYARKHADNAAKHSL
CNAVNRDIFGLQAYADQVDLTIKIPLANGHEDETHALHLAVRSVORTSLHI VDFLV
QNSNLDKTQKGTALHYCCLTQNAECLKLLRGKASIEIANESGETPLDIKRLKH
EHCBELTQALSGFNSHVHYEWRLHEDSDDDMDKELQPSNRRDRPISFY
OLGNSQALSVLADANLAKKQRAFPESILQNETYGALLSGSPPAOPAPSTT
SAPPLPRNVKQVTOASSANTLWKTNSVSVDSGSRORSSDPBAVHPPLPLRVSTN
PLPTPPPPVAKTSVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKPPAGADK
STPLTNKGQPRGPDLSATEALGPLSNWNLQPPAPMPKRSQATKLKPRVKALYNV
ADNPDELTFSEGDVIIVDGEEDQWMIHIDGDFGRGAFFVSVFHTAD"

```

## gene

## CDS

Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle	160
Db	761	GATCTGAAAAAGCCCTTTTGTATAAGACTTGAAGAGACTATGAACAAAAAATACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLysGlyMetIleArgThrGluLysSerGly	180
Db	821	GAAGAAGGAAAAAGGAAACAGCCCAAGCTCCATCGGATGATTCGAGCTGAATTAAGCGGA	880
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	881	CGCGAAATTCGGAAGAGATGGAAGAGAGAGGCGCTCTTCCAGCTACAGATGTGCGAG	940
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	941	TATCTGCTGAAGGTCACGAAGAAATCAAGATTAAAAAGGGAGTAGATTTACTTTCAGAAATCTG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAAAATCTTTCATGCCAATGCATTTTTTTCAGGATGGACTCAAGCCGCTGGNAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAAACTCTTCATTGAAACGCTGCTACGGATCTTACACAGATCAAAACAGGCCCAGGAT	1120
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuValGlu	280
Db	1121	GAAGAAAGAGAGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCGAGTTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGAGGACTCCCAAAATTCGTACAGGACACAGCTTATAGCTTACATCAGCCTCAGGA	1240
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AACAAGAACATGGGACCGAGCGGACGCGAGCTCTTACAAGAAAGAGTGAAGGATCCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAAGGAAATGTTTCAAGTAAAAAATGGTTTTCTGACCATATCCATGGT	1360
Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGGCTCTCGCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCT	1420
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAGAAGAAGTGTCTTGACCTTATTTTCATCATGACAGAACTTACCACCTTTCAAGCTGAA	1480
Qy	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluLysAlaLeu	400
Db	1481	GATGAAAGGAAATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAGAAAGAGCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1541	AACAATGCATTTAAGGGGGATGCAATACTGGAGAAAATAACATCGTCCAAGAACTGACA	1600
Qy	421	LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGCTGTGCTGTGACTGTGGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1661	CGCGCAGATCTCATCGCTTTCCACCAACCTGGGCACTCCGAGCTGCATCGATGTGTTCC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVal	480
Db	1721	GGATCCACCGAGAGCTGGGGGTTCAATATTTCCAG-GATGCAGTCCCTGACCTTAGATGT	1779
Qy	480	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1780	ACTGGGAACATCTGAGCTGCTCGCAAGAAATATTTGGGAATATTTGGGAATCGAGCTTTAATGAGAT	1839

## ORIGIN

Pred. No.:	0	Length:	5711
Score:	717.00	Matches:	1003
Percent Similarity:	99.41%	Conservative:	0
Best Local Similarity:	99.41%	Mismatches:	3
Query Match:	71.27%	Indels:	6
DB:	9	Gaps:	0
US-09-914-042-1 (1-1006) x AB007860 (1-5711)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla	20
Db	341	ATGCGGACACAGATCTCCGTTGCGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG	400
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	401	CCCACGGCTCCAGCTTCAACACCCGACGCGCAGTGCAGGAACTGTCGGCCCATC	460
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	461	GAGAGGGCTTTGGACCTGGACCGGATGTTCTTTACAAAATAGAGAAATCCGTGAAGCA	520
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu	80
Db	521	ATCAACAGCTCTGGGCTGGCTCACGTGCAAAATGAAGAGCAGTACACCCAGGCTCTGGAG	580
Qy	81	LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys	100
Db	581	AAGTTTGGCGCAACTGTGTATGACAGAGATGACCCAGATTTTAGGAAGTGCCTTCCCTGAAG	640
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	641	TTCTCAGTGTTCNAAGAGAGTTGACAGCATCTTTCAAAAACCTGATTTCAGAAATATGAAC	700
Qy	121	AsnIleLeuSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	701	AACATAATCTCTCTCCCTTTGACAGATTTGCTGAAGGGGGACCTGAAAGGAGTGAAGGG	760

QY	500	eMetGluCysCysLeuProAlaAspSerValLysProAsnProGlySerAspMetAs	520
Db	1840	CATGGAAATGTTCCTACCAAGCTGAGGACTCAGTCAAAACCAACCCAGGAGGACATGAA	1899
QY	520	nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHi	540
Db	1900	TCCAGAAGGACTACATCACAGCCAAAGTACATCGAGAGGAGATACGCAAGGAAGCA	1959
QY	540	sAlaAspAsnAlaAlaLysLeuHiSerLeuCysGluAlaValLysThrArgAspIlePh	560
Db	1960	CGCGGATAACCGCGGCGAAGCTTCACAGTCTTTTCGAGGCGGTCAAAACGAGAGATATTTT	2019
QY	560	eclyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs	580
Db	2020	TGGATTGCTCCAAAGCTTATGCTGATGGTGTGGATCTTACGGAAAAATCCCACTGCGCAA	2079
QY	580	nGlyHisGluProAspGluThrAlaLeuHiSerLeuAlaValArgSerValAspArgThrSe	600
Db	2080	CGGACATGAGCGCGATGAACCGGCTCCACCTTGAGTTCAGTTCAGTTCGATCGAACCTTC	2139
QY	600	rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy	620
Db	2140	TCTTCACATTGTAGACTTTTATGTTTCAGAACAGTGGGAACCTGGATAAACACAGAGGAA	2199
QY	620	sGlySerThrAlaLeuHiSerTyrCysCysLeuThrAspAsnAlaGluCysLeuLeuLe	640
Db	2200	AGGAGCACAGCCCTGCACTACTGCTGCTGACCAATGCGAGTGCCTCAAGTTGCT	2259
QY	640	uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIl	660
Db	2260	CCTCGGGGGAAGGCTTCATCGAGTAGCAACAGAGTCAGAGAGACTCCGCTGGACAT	2319
QY	660	eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr	680
Db	2320	TGCCAAGCGCTCAAGCAGCAGCACTGTGAGGAGCTGTGACCCCAAGCTTATCTGGAG	2379
QY	680	gPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGl	700
Db	2380	ATTTAAATCTCACGTTTCAGTTGAATATGATGGCGACTACTTCCACGAAGACCTGGATGA	2439
QY	700	uSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgP	720
Db	2440	AAAGTATGACGATCGATGGATGAGAAATGCGGCCAGTCC-CAACCGCGGGGAAGACCGGC	2498
QY	720	roIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaA	740
Db	2498	CCATCAGCTTCTACAGCTGGCTCCACCAAGCTTCAGTCTAACCTGTATCTTTGGCA	2558
QY	740	rgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnA	760
Db	2558	GAGATGCTGCAAACTTGCAGGAGAGAGCAGAGGGCTTTTCATGCCAGCATCTTTGCAGA	2618
QY	760	ngLuthTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProS	780
Db	2618	ATGAGCTTACGAGCCCTCTGAGTGAGCGCCCACTCCGCGCCAGGCTTCGAGCCCA	2678
QY	780	erThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS	800
Db	2678	GCACACACAGCGCCCCCGCTTCTCCAGGAATGTTGGCAAGTTTCAGACAGCTCTCT	2738
QY	800	eAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS	820
Db	2738	CTGCTAACACCTGTGGAAGCAAACTCTGTAAGTGTGAGCGGTGGAAGCGGACGAT	2798
QY	820	erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA	840
Db	2798	CTTCGTCAGATCCGCGAGCTGCCATCCACGCTGCCCTCTTCGCGTGACATCTACCA	2858
QY	840	snProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaL	860
Db	2858	ATCCCTGACCCCAAGCGGCCCCCAGCCGTTGCCAAGACGCCAGCGGTATGGAGGCT	2918
QY	860	eSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuP	880

Db	2919	TTAGCCAGCCGAGCAAGGCTGCCCGCCTGGGATCTCAGATCAGGCCACCTCTGC	2978
QY	880	toProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer	899
Db	2978	CCCCACAGCCGCCAGCGCTCCCGCAGAGAGACCTGCGCGGGGGT-GACAAAGTCC	3037
QY	900	ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla	919
Db	3037	ACCCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGAAGCT	3097
QY	920	LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer	939
Db	3097	CTGGGTCTCTGTCTCAATGCTATGCTGTCGAGCCCTGTCACCCATGCTTAGGAAGTGC	3157
QY	940	GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValaAlaAspAsn	959
Db	3157	CAGGACACCAAGTTGACGCTTAAGCGGTGAAGCGCTCTATTAATCTGTGTGGCTGACAC	3217
QY	960	ProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGln	979
Db	3217	CCCGATGAGTCCACCTTCTCCGAGGGGATGTGATCATCTGTCGAGCGGAGGAGACCCAG	3277
QY	980	GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer	999
Db	3277	GAGTGTGGATTGGCCACATTCATGATGAGATCTGTGTCGAAAGGCGCATTCCTCGGTGTC	3337
QY	1000	PheValHisPheIleAlaAsp	1006
Db	3337	TTTGTGCACTTATCGCTGAC	3358
RESULT 2			
LOCUS	BC063308	5544 bp	mRNA linear PRI 23-DEC-2003
DEFINITION	Homo sapiens development and differentiation enhancing factor 2, mRNA (cDNA clone MGC:71731 IMAGE:30345039), complete cds.		
ACCESSION	BC063308		
VERSION	BC063308.1		
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 5544)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toehiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullan, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.C., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Bonfield, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywicki, M.I., Skalka, U., Smal, J., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 5544)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. Stefan Hansson  
cDNA Library Preparation: Michael Brownstein / Ted Usdin  
Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@axil.stanford.edu](mailto:mcd@axil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 133 Row: h Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4502248.

# FEATURES

## source

1..5544  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:71731 IMAGE:30345039"  
/tissue\_type="Placenta, normal"  
/clone\_lib="NIH\_MGC\_147"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
1..5544  
/gene="DDEF2"  
/note="synonyms: KIAA0400, PAP, SHAG1, PAG3, Pap-alpha"  
/db\_xref="LocusID:8853"  
/db\_xref="MIM:603817"  
294..3179  
/gene="DDEF2"  
/codon\_start=1  
/product="DDEF2 protein"  
/protein\_id="AAH63308.1"  
/db\_xref="GI:38648787"  
/db\_xref="LocusID:8853"  
/db\_xref="MIM:603817"

## gene

## CDS

/translation="MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATEEAL  
DVRNVLKMKSVKAINSGLAHVENEEQVTALEKEGNGRVCDDPDGLGSAFLKES  
VFTKELTALFNLIQNMNNIISFPLSLKGLDKGDKLKKPDKAKWDYETKITKI  
EKEKHEAKLHGMIRTEISGAIEAEMEKEKRRFFQMCQYELLKVKNEIKIKKGVLDLQ  
NLIKYFAQCNFQDGLKAVESLKPSETLSTDLTIKQADERRQLIQRDLKSA  
LQVEKQDSQIRQSTAYSLHQPQGNKEHGTNRNGLSKYKSGIRKWKQKCSVKNQF  
LTISHGTANRPAPKLLNLTQVNTNPEKKCFDLISHDRTYHFOADEQEQIWMNVL  
QNSKEALNNAFKGDDNTGNNIYQELTKELISEVQMTGNDVCCDGPAPETWLSN  
LGILTCISGHRRLGVLHYHSMQSLTDVIGTSELLAKNIGNAGFNEIWECCLPAP  
DSVRKPNQSDNWKADYITAKYIERRYARKHADNAKLSLCEAVKTRDIFGLLQAY  
ADGVDLTEKIPLANGHEPDETHALHVAHSVDRSLHIVDFVQNSGNDKQTKGKSTA  
LHYCLCTONAEKLLLRKASIEIANESGETPLDIKRLKHECEELLTQALSGRFN  
SHVHVEYRLLHLEDSDDDMDKLOPSPNRRDRPISFYOLGSLNQLQSNVSLAR  
DAANLAKQKRAFMSIILQNETYQALLSGSPPPAPAPSTTSAPPLPRNVKQDPLT  
FTPPPVAKTSVMEALSQSPKAPPGISQIRPPLPQPPSRRLQKPKAPGADKSTP  
LTNKKQRPGLVDLSATEALGPLSNAMVLOPPAPMPKQSOATKLKPRKVALYNCVADN  
PDELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAPFVSFVHFIAID"  
1209..1490  
/gene="DDEF2"  
/note="PH; Region: Pleckstrin homology domain"  
/db\_xref="CDD:smart00233"  
1554..1898  
/gene="DDEF2"  
/note="ArfGap; Region: Putative GTPase activating protein  
for Arf. Putative zinc fingers with GTPase activating  
proteins (GAPs) towards the small GTPase, Arf. The GAP of  
Arf1 stimulates GTPase hydrolysis for ARF1 but not ARFs"  
/db\_xref="CDD:pfam01412"  
2052..2312  
/gene="DDEF2"  
/note="ANK; Region: ankyrin repeats"

## misc\_feature

## misc\_feature

1073 TCAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAATATCCCGATTGCAGGTGA 1132

misc\_feature  
2991..3170  
/db\_xref="CDD:cd00204"  
/gene="DDEF2"  
/note="SH3; Region: Src homology 3 domains"  
/db\_xref="CDD:smart00326"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 5544  
Score: 525.00 Matches: 790  
Percent Similarity: 99.00% Conservative: 0  
Best Local Similarity: 99.00% Mismatches: 4  
Query Match: 52.19% Indels: 8  
Gap: 9

US-09-914-042-1 (1-1006) x BC063308 (1-5544)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLeuAla 20  
Db 294 ATGCCGACCCAGATCTCCGTGTCGGAATTCGTGCCGAGACCCCATGAGGACTCAAGGCG 353  
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
Db 354 CCCACGGCTCCAGCTTCACCACCCGACGCGCAGTCCCGAAACACCTGTGGCGCGCATC 413  
Qy 41 GluGluAlaLeuAspValAspAtqMetValLeuTyrLysMetLysLysSerValLeuAla 60  
Db 414 GAGGAGGCTTTGGACGTGGACCGGATGTTCTTTACAAATGAAGAAATCCGTGAAGCA 473  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluIntyrThrGlnAlaLeuGlu 80  
Db 474 ATCAACAGCTCTGGCTGGCTCACGTGAAATCAAGAGCAGTACACCCAGGCTCTGGAG 533  
Qy 81 LysPheGlyGlyAsnCys-ValCysArgAspAspProAspLeuGlySerAlaPheLeu 100  
Db 534 AAGTTTGGCGCAACCG-TGTATGCAGAGATGACCCAGATTTAGGAAAGTGCGTTCCTGAA 592  
Qy 100 sPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAs 120  
Db 593 GTTCTCAGTGTTTACAAGGAGTGTGACAGCACTTTTCAAAACCTGTATTCAGATATGAA 652  
Qy 120 nAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLeuG 140  
Db 653 CAACATAATCTCTCCCTTTGGACAGTTTCTCTGAAGGGGACCTGAAAGGAGTGAAGG 712  
Qy 140 YAspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysI 160  
Db 713 GGATCTGAAAAGGCTTTTGATAAGCTTGAAGGAGCTATGAAACAAATAAACCAAGAT 772  
Qy 160 eGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerG 180  
Db 773 AGAAAGAGGAAAGAGAACACCGCAAGCTCCATGGGATGATTCCGACTGAATAAGCGG 832  
Qy 180 YAlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysG 200  
Db 833 AGCGAAATTCGCCAAGAGATGGAAGAGGAGGCGCTTCTTCCAGCTACAGATGTGCGA 892  
Qy 200 uTyrLeuLysValLeuGluLysIleLysIleLysGlyValAspLeuLeuGlnAsnLe 220  
Db 893 GTATCTGCTGAAGGTCAACGAAATCAAGATTAAAGAGGAGTAGATTACTTCAGAAATCT 952  
Qy 220 uIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSe 240  
Db 953 GATCAATATCTTTTATGCCCAATGCAATTTTTTTTCAGGATGAGCTCAAGGCGGTGAAG 1012  
Qy 240 rLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrLysGlnAlaGlnAs 260  
Db 1013 CCTCAAACTTCCATTGAAACGCTGTCTACGATCTTCCACAGATCAAAACAGGCGCCAGGA 1072  
Qy 260 pGluGluArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValG 280  
Db 1073 TCAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAATATCCCGATTGCAGGTGA 1132

```
Qy 280 uGlnlySGluAspSerGlnIleAtrGlnSerThrAlaTyrsSerLeuHisGlnProGlnG1 300
Db 1133 ACAGAAAGAGAGCTCCCAAAATTCGTGACAGACAGCTTATAGCTTACATCAGCTCAGGG 1192
Qy 300 yAsnlySGluHisGlyThrGluAtrGlnSerLeuTyrsLeuTyrsSerLeuHisGlnProGlnG1 320
Db 1193 AAACNAGGAAATGGGACCGGAGCGGACGCTCTACAGAAAGAGTACGGGATCCG 1252
Qy 320 gLysValTrpGlnlyAspGlyCysSerVallyAsnGlyPheLeuThrIleSerHisG1 340
Db 1253 AAAAGTGTGGCAGAAAGCAATGTTTCAGTTAAATAATGTTTCTGACCATATCCCATGG 1312
Qy 340 yThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnVallyThrAsnPr 360
Db 1313 TACCGCTAACCGGCTCTCGAAAGCTCAACCTGCTAACTGCCAGGTGAAGCAACCC 1372
Qy 360 oGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrsPheGlnAlaG1 380
Db 1373 TGAGGAGAAAGTGTGTTGACCTTATTTTCAATGACAGAACTTACCACTTTCAAGCTGA 1432
Qy 380 uAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLe 400
Db 1433 AGATGAACAGGAATGTCAAATATGATGTCTGTGTCGCAAAATAGCAAGAAAGAGCTTT 1492
Qy 400 uAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuTh 420
Db 1493 AAACAATGCATTTAAGGGGGATGACATATCTGGAGAAATATACATCGTCCAAAGAACTGAC 1552
Qy 420 rLysGluIleLeuSerGluValGlnAtrMetThrGlyAsnAspValCysCysAspCysG1 440
Db 1553 AAAGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATGACGCTCTGTGACTGTGG 1612
Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
Db 1613 GGGCCAGATCTACATGCTTTCCCAACCTGGGCATCTCGACCTGATCGAGTGTTC 1672
Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrsSerPro-MetGlnSerLeuThrLeuAspV 480
Db 1673 CGGAATCACCAGAGAGCTGGGGTTCATTATTCAG-GATCAGTCCCTGACCTTAGATG 1731
Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500
Db 1732 TACTGGGAACATCTGAGCTGTCTGCGCAAGATATTTGGGAATGCGAGCTTTTAAATGAGA 1791
Qy 500 leMetGluCysCysLeuProAlaGluAspSerVallyAsnProGlySerAspMetA 520
Db 1792 TCATGGATGTTCCTACAGCTGAGGACTCAGTCNAACCCAAACCCAGCAGCAGCATGA 1851
Qy 520 snAlaArgLysAspTyrlleThrAlaLysTyrlleGluArgArgTyralaArgLysLysH 540
Db 1852 ATGCAAGAAAGGACTTACATCAGCAAGTACATCGAGAGGAGATACGCAAGGAAGAAGC 1911
Qy 540 isAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaVallyThrArgAspIleP 560
Db 1912 ACGCGGATAACGGCGGAGCTTACAGCTCTTCGAGGCGCTCAAAACGAGAGATATTT 1971
Qy 560 heGlyLeuLeuGlnAlaTyrsAlaAspGlyValAspLeuThrGluLysIleProLeuAlaA 580
Db 1972 TTGATGTCTCCAGCTTATGCTGATGTGTGATCTTACGGAATAATCCCATGCGCA 2031
Qy 580 snGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrS 600
Db 2032 ACGGACATGAGCGGATGAAACGGCCCTCCACCTTGCACTGAGTCCGATCGATCGAACCT 2091
Qy 600 erLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyL 620
Db 2092 CTCTTCATATGTAGACTTTTGTAGTTCAGAACAGTGGGAACCTGGATAAACAGACAGGA 2151
Qy 620 ysGlySerThrAlaLeuHisTyrsCysLeuThrAspAsnAlaGluCysLeuLysLeuL 640
Db 2152 AAGCAGCAGACAGCTTACTGCTGCTGACCGACATGCCAGTGCCTCAAGTTGC 2211
Qy 640 euLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspI 660
```

```
Db 2212 TCCTGCGGGGAAAGCGCTCCATCGAGATAGCAAAACGAGTCAGGAGACTCCGCTGGACA 2271
Qy 660 leAlaLysAtrGLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyA 680
Db 2272 TTGCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGCTGACCCCAAGCCTTATCTGAA 2331
Qy 680 rgPheAsnSerHisValHisValGluTyrgluTyrgluLeuHisGluAspLeuAspG 700
Db 2332 GATTAAATTCACGTTTCAGCTTCAATATGATGATGATGATGATGATGATGATGATGATG 2391
Qy 700 luSerAspAspAspMetAspGluLysLeuGlnProSerGlu-AsnAtrArgGluAspArg 719
Db 2392 AAAGTGAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2450
Qy 720 ProIleSerPheTyrglnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAla 739
Db 2451 CCCATCAGCTTCTACAGCTGGGCTCCACAGCTTCTAGTCTAAGCTGTATCTTTGGCC 2510
Qy 740 ArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuG1 759
Db 2511 AGAGATGCTGCAAACTTGCACAGGA-CAAGCAGAGGGCTTTTCATGCCAGCATCTTGCA 2569
Qy 759 nasnGluThrTyrglyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaPr 779
Db 2570 GAATGAGACTTACGAGAGCCCTCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGCAGCCC 2629
Qy 779 oSerThrThrSerAlaProProLeuProProAtrGlnValGlyLys 794
Db 2630 CAGCACCCAGCCGCCCGCTTCTTCCACGGAATGTTGGCAAA 2675

RESULT 3
LOCUS I74317 4486 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 15 from patent US 5688641.
ACCESSION I74317
VERSION I74317.1 GI:3010458
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4486)
AUTHORS Sager,R., Zou,Z., Lee,S.Whan. and Tomasetto,C.Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 15 18-NOV-1997;
FEATURES
source 1..4486
/mol_type="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2,11e-292 Length: 4486
Score: 297.00 Matches: 701
Percent Similarity: 98.59% Conservative: 0
Best Local Similarity: 98.59% Mismatches: 4
Query Match: 29.52% Indels: 10
DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x I74317 (1-4486)
Qy 302 LysGluHisGlyThrGluAtrGlnSerLeuTyrsLysSerAspGlyIleArgLys 321
Db 1 AAGGAACATGGGACCGAGCGGACGCGAGCTCTTACAAAGAGAGTACGGGATCCGAAA 60
Qy 322 ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341
Db 61 GTGTGGCAGAAAAGAAATGTTTCAATGATGATGATGATGATGATGATGATGATGATG 120
Qy 342 AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnVallyThrAsnProG1 361
Db 121 GCTAACCGGGGCTCTCTCAAGGCTCAACCTGCTTAACTGCTGAGGTGAAGCAACCTTGA 180
```

Qy 361 uGluLysLysCysPheAspLeuLeuSerHisAspArgThrTyrHisPheGlnAlaGluAs 381  
Db 181 GGAGNAGAGTGTCTTGACCTTATTTACATGACAGAACTTACCACCTTTCAGCTGGAAGA 240  
Qy 381 pGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAs 401  
Db 241 TGAACAGGAATGTCAAAATATGATGTCTGTGTCGCAAAATAGCAAGAAGACTTTAA 300  
Qy 401 nasnAlaPheLysGlyAspAspAsnThrGly<sup>3</sup>HisAsnIleValGlnGluLeuThrLys 421  
Db 301 CAATGCATTTAAGGGGGATGACAACTACTGGAGAAATAACTCGTCCAAAGAACTGACAAA 360  
Qy 421 sGluIleLeuSerGluValGlnAtqMetThrGlyAsnAspValCysCysAspCysGlyAl 441  
Db 361 GGAGATCATCTCAGNAGTGCAGAGATGACGGGCAATGCTGCTGTGCTGTGCTGGGGC 420  
Qy 441 aProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGl 461  
Db 421 GCCAGATCCTACATGGCTTTCACCAACCTGGGCATCCTGACCTGCATCGAGTGTTCGG 480  
Qy 461 yIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValL 481  
Db 481 AATCCACGAGAGCTGGGGGTTCATTATTTCAG-GATGCAGTCCCTGACCTTAGATGTAC 539  
Qy 481 euGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleM 501  
Db 540 TGGGNACATCTGAGCTGTCTCGCAAGAATATTGGGAATGCGAGCTTTAATGAGATCA 599  
Qy 501 etGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnA 521  
Db 600 TGGAAATGTGCTACCACTGAGGACTCAGTCAAAACCCCAACCCAGGACGACATGAATG 659  
Qy 521 laArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisA 541  
Db 660 CAAGAAAGGACTACATCACCAAGTACATCGAGAGAGATACGCAAGGAAAGACG 719  
Qy 541 laAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheG 561  
Db 720 CGGATTAACCGCGGAGGCTTACAGTCTTTGCGAGGCGCTCAAAACGAGAGATATTTTG 779  
Qy 561 lyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleLeuProLeuAlaAnG 581  
Db 780 GATTGCTCAAAGCTTATGCTGATGTGTGGATCTTACGGAAAAAATCCACTGGCCAAAG 839  
Qy 581 lyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerL 601  
Db 840 GACATGACCGGATGAACCGCCCTCCACCTTGCACTGAGTCCGATCGGATCGAACCTCTC 899  
Qy 601 euHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysG 621  
Db 900 TTCACATTTGAGACTTTTATGTTCAAGACAGTGGGAACCTGGATAAACAGACAGGGAAG 959  
Qy 621 lySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuL 641  
Db 960 GCAGCACAGCCCTGCACCTACTGCTGCTGACCGACAATGCCAGTGCCTCAAGTTGCTCC 1019  
Qy 641 euArgGlyValAsaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleA 661  
Db 1020 TGGCGGGGAAGGCTCCATCGAGATAGCAAAATGAGTCAAGGAGAGACTCCGCTGGCAATG 1079  
Qy 661 laLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgP 681  
Db 1080 CCAAGCGCTCAAGACAGGACACTGTGAGGAGCTGCTGACCCAGCCCTTATCTGGAGAT 1139  
Qy 681 heAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluS 701  
Db 1140 TTAATTCCTCAGCTTCACGTTGAATATGAATGCGGACTACTCCACGAAGACCTGGATGAA 1199  
Qy 701 erAspAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgPro 720  
Db 1200 GTGATGACGATGATGAGAAATTCAGCCAGTCC-CAACCGCGGGAAGACCGGCC 1258  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740

Db 1259 ATCAGCTTCTACAGCTGGGCTCCACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 1318  
Qy 741 AspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuGlnAs 760  
Db 1319 GATGCTCAAAACCTTGCCAAAGGA-CAAGCAGAGGGCTTTTCATGCCCAGCATCTTGCAGAA 1377  
Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSe 780  
Db 1378 TGAGACTTTAGCGAGCCCTCTGAGTGGCAGCCCACTCCCGCCAGGCTGCGAGCCCCAG 1437  
Qy 780 rThrThrSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS 800  
Db 1438 CACCACAGCGCCCGCCCGCTTCTCCACGGAATGTTGGCAAGTTTCAGACAGCTCTCT 1497  
Qy 800 erAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820  
Db 1498 CTGCTAAACACCTGTGGAAGACAAACTCTGTAAGTGGACGCGTGGGAAGCGCGCAGCAT 1557  
Qy 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA 840  
Db 1558 CTTGCTCAGATCCCGCAGCTGTCCATCCACGCTGCCCGCTTCTCGGCTGACATCTACCA 1617  
Qy 840 snProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaL 860  
Db 1618 ATCCCTTGACCCCAACCGCCCGCTTCCAGAGACGCCAGCGGTAATGGAGCCT 1677  
Qy 860 euSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuP 880  
Db 1678 TGAGCCAGCGGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCCAGCTCTGC 1737  
Qy 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899  
Db 1738 CCCACAGCGCCCGCAGCGCTCCCGCAGAAAGAAAGCTGCGCCAGGGGCT-GACNAGTCC 1796  
Qy 900 ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla 919  
Db 1797 ACCCACTGACCAACAAGGCCAACCCGAGAGGACTGTGGATCTCTCTGCAACCGNAGCT 1856  
Qy 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939  
Db 1857 CTGGGCTCTCTGTCCAATGCTATGCTCTGCGAGCCCCCTGCACCCATGCGCTTAGGAAGTCG 1916  
Qy 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959  
Db 1917 CAGCAACCAAGTTGAAGCCCTAAGCGGCTGAAGAGCTCTATAACTGTGTGGCTTGACAA 1976  
Qy 960 ProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGln 979  
Db 1977 CCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGACGGGAGGAGGACCAG 2036  
Qy 980 GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999  
Db 2037 GAGTGTGGATTTGGCCACATTTGATGGAGATCTCTGTCGCAAGGCGCATTTCCCGGTGTCA 2096  
Qy 1000 PheValHisPheIleAlaAsp 1006  
Db 2097 TTTGTGCATTTATCGCTGAC 2117  
RESULT 4  
LOCUS I74314 4328 bp DNA linear PAT 03-APR-1998  
DEFINITION Sequence 12 from patent US 5688641.  
ACCESSION I74314  
VERSION I74314.1 GI:3010455  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4328)  
AUTHORS Sager R., Zou Z., Lee S., Whan, and Tomasetto, C. Laure.  
TITLE Cancer diagnosis using nucleic acid hybridization  
JOURNAL Patent: US 5688641-A 12 18-NOV-1997;

FEATURES  
source Location/Qualifiers  
1..4328  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.55e-251 Length: 4328  
Score: 257.00 Matches: 648  
Percent Similarity: 98.63% Conservative: 0  
Best Local Similarity: 98.63% Mismatches: 4  
Query Match: 25.55% Indels: 9  
DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x 174314 (1-4328)

QY 355 GlnValLysThrAsnProGluGluLysLysCysPheAspLeuLeuSerHisAspArgThr 374  
DB 3 CAGGTGAAGACCAACCTTGAGGAGAGAGTGTCTTGACCTTATTTTCATGACAGAACT 62  
QY 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsn 394  
DB 63 TACCACCTTTCAAGCTGAAGTGAACAGGAAATGTCAAATATGATGCTGTGCTGCAAAAT 122  
QY 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn 414  
DB 123 AGCAAGAGAGAGCTTTAAACAATGCATTTAAGGGGGATGACAATCTCGAGAAATAAC 182  
QY 415 IleValGlnGluThrLysGluIleLeuSerGluValGlnAcmMetThrClyAsnAsp 434  
DB 193 ATGCTGCAAGAACTGACAAAGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATGAC 242  
QY 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu 454  
DB 243 GTCTGCTGACTGTGGGGGCGCCAGATCTACATGCTTCCACCAACTGGGCATCTG 302  
QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1 474  
DB 303 ACCTGCATCGAGTGTTCGGGAATCACCAGAGAGTGGGGTTCATTATTCAG-GATGCA 361  
QY 474 nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs 494  
DB 362 GTCCCTGACCTTAGATGACTGGGAACATCTGAGCTGCTGCTCGCAAGAATATTGGGAA 421  
QY 494 nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs 514  
DB 422 TGCAGCTTTAATGAGATCATGGAATGTTCCTACAGCTGAGGACTCAGTCAAAACCAA 481  
QY 514 nProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgAr 534  
DB 482 CCCAGGCGAGCATGAATGCAAGAAAGGACTTACATCATCAGCCAAAGTACATCGAGAGGAG 541  
QY 534 gTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVa 554  
DB 542 ATACCAAGGAAGAGACGCGGATTAACGCGCGAAGCTTCACAGCTTTTCGAGGGCGGT 601  
QY 554 lLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrG1 574  
DB 602 CAACACGAGAGANTATTTTGGATGCTCCAGCTTATGCTGATGCTGATGCTGATCTACGGA 661  
QY 574 uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr 594  
DB 662 AAAATCCCACTGGCCCAACGAGCATGAGCCGATGAACGGCCCTCCACCTTCAGAGTCA 721  
QY 594 gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe 614  
DB 722 ATCCGTGATGCAACTCTCTTCACTATGTAGACTTTTATGTTCAAGACAGTGGGAACCT 781  
QY 614 uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAl 634  
DB 782 GGATAAACAGACAGGGAAGGACGACAGCCCTGCACTACTGCTGCTGCTGCTGCTGCTG 841  
QY 634 aGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerG1 654

DB 842 CGAGTGCCTCAAGTTGCTCTCGCGGGGAAGCCCTCCATCGAGATAGCAAAATAGACTCAGG 901  
QY 654 yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlyGluLeuLeuTh 674  
DB 902 AGAGACTCCGTGGACATTTGCCAAGCGCTCAGACACGAGCACTGTGAGGAGCTCTGAC 961  
QY 674 rGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLe 694  
DB 962 CCAAGCCTTATCTCGAAGATTTAATCTTCACGTTCCAGCTTGAATATGAATGCGGACTACT 1021  
QY 694 uHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu-A 714  
DB 1022 CCACGAAGACCTGGATGAAGTGAATGACGACATGATGAGAAATTTGCAGCCAGTCC-CA 1080  
QY 714 snArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerA 734  
DB 1081 ACCGGGGGAGACCGGCCCATCAGCTTCTACAGCTGGGTCCAAACAGCTTCACTCTA 1140  
QY 734 snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe 753  
DB 1141 ACCTGTATCTTTGGCCAGAGATGCTGCAAACTTTGCCAAGGA-CAAGCAGAGGGCTTTC 1199  
QY 754 MetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProPro 773  
DB 1200 ATGCCAGCATCTTTCAGAATGAGACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCC 1259  
QY 774 AlaGlnProAlaAlaProSerThrSerAla-ProProLeuProProArgAsnValG1 793  
DB 1260 GCCCAGCCTGCGGCCCCCGCAGCACCCAGCCGCCCCCGCTTCTCCCGAATGTTGG 1319  
QY 793 yLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAs 813  
DB 1320 CAAAGTTTCAGACAGCTCTCTCTGCTAACACCTGTGGAAGACAAACTCTGTAAGTGTGA 1379  
QY 813 pGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProPr 833  
DB 1380 CGGTGGAAGCGGAGCGATCTTCGTGAGTCCGCGAGCTGTCCATCCACCGCTCCGCC 1439  
QY 833 oLeuArgValThrSerThrAsnProLeuThrProThrProProProValAlaLysTh 853  
DB 1440 TCTTGGGTGACATCTACCAATCCCTGACCCCGCCAGCCGCCCGCCCGCTTGCACAGAC 1499  
QY 853 rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProGlyIleSerG1 873  
DB 1500 GCCCAGCGTAATGAAGCCTTTAGCCAGCCGAGCAAGCCTGCCCGCTGGGATCTCACA 1559  
QY 873 nileArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAl 893  
DB 1560 GATCAGGCCCCCACTCTGCCCCCAGCCGCCCGCCCTCCCGCAGAAAGAGCCTGC 1619  
QY 893 aPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA 913  
DB 1620 GCCAGGGGCT-GACAAGTCCACCCCTGACCAAAAGGCGCAACCCGAGAGACCTGTGG 1678  
QY 913 spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA 933  
DB 1679 ATCTCTGCAACGAGAGCTCTGGGTCTCTCTGCTCCAATGCTATGCTCTGCGAGCCCTG 1738  
QY 933 laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT 953  
DB 1739 CACCATGCTCTAGGAAGTCCGAGGCAACCAAGTTGAAGCCTTAAGCGGTGAAAGCGCTCT 1798  
QY 953 yzAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleLev 973  
DB 1799 ATAACTGTGTGGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCG 1858  
QY 973 alAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgL 993  
DB 1859 TGGACGGGAGGAGGACGAGGATGTGGATTTGGCCACATTTGATGGAGATCTCTGTCGCA 1918  
QY 993 ySgLyAlaPheProValSerPheValHisPheIleAlaAsp 1006



```
Db      1919 AAGCGCATTCGGGTGTCATTGTGTCACCTTTATCGCTGAC 1959
RESULT 5
AKI72944 LOCUS AKI72944 4713 bp mRNA linear ROD 28-JUL-2004
DEFINITION Mus musculus mRNA for mKIAA0400 protein.
ACCESSION AKI72944
VERSION AKI72944.1 GI:50510472
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Koseki,H.,
Hiraoka,S., Saga,Y., Seino,S., Nishimura,M., Kaisho,T., Hoshino,K.,
Kitamura,H., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the Coding Sequences of Mouse Homologues of KIAA
Gene: IV. The Complete Nucleotide Sequences of 500 Mouse
KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences
of cDNA Clones Randomly Sampled from Size-Fractionated Libraries
JOURNAL DNA Res. 11, 205-218 (2004)
REFERENCE 2 (bases 1 to 4713)
AUTHORS Okazaki,N., Kikuno,R.F., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2004) Hiraashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing: Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
source
1..4713
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mfj04247"
/note="vector:modified pBC SK+"
1..4713
/gene="mKIAA0400"
<65..2977
/gene="mKIAA0400"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence="not experimental"
/product="mKIAA0400 protein"
/protein_id="BAD3222.1"
/db_xref="GI:50510473"
/translation="GGGSGPRGVAAMPDQISVSEFVAETHEDYKAPTASSFTTRTAQ
CRNTVAIEALDVRMVLVYRMKKSVAKINISGLAHVENEQYQTALEKFGNCVCVR
DPDLGSALEAFSVFTKELTALFKNLIQNNMIISPLDSLKGLKGKGLKXPKFDK
AWKDETITKIEKEKHAHLGMIRTEISGAETABEMEKERRFFQLOMCVEYLLKVN
EIKYQRLDKLQNLKYKHAQCNFQDGLKAVESIKPSIETSLDTHLTKQAQDEER
QLIQRDLKSLAQVEKESRDSQLRSTAYSLHPQGNKEHGTENGNLYKKSQGI
RKVMQKCSKWSGFLNLTISHTANRPAPKLNLTQCVNTNPEEKCFDLISHDRTHF
QAEDEQECQIWMVQLNSKEALNNAFKGDDNTGNNIVQELTEIISSEVQMTGNDV
CDDCAPPTWNLGLTICTEGSIHREGLVHYSRMQSLTLDVLGTSLLAKNLG
NAGFNEIMECCPLSPDPKVPNGSDMIARKDYITAKVMRYARKKHADTAALKSLC
EAVKTRDIFGLQAVDGVLDTEKIPLANGHEPDETALHLAVRSVDTPSLHVDPLVQ
NSGNDLKQTGSLSTALHYCCLTDNNAECLLLRLGKASIEIANSEGETPLDIARLKH
HCEBLLTQALSGRNFVHVHVEYERLLHLDDESDDDVDEKLQSPNRRDRPVSYFQ
LGSSQFQNAVSLARDTANLTKQRFQGPSILQNETYGAILSGSPSPSSQIPFSTTS
APPILPRNVKQDLPTTPPPVAKTSGTLEAMNPKSKSQPGTSQSKPPLPPPPRVR
LPQKPKASGTDKPTPLTNKGQPRGPPEASGPLSNMALQPPAPMPRKSDATKSKPKRVK
ALYNCVADNPDELTFSEGDVIVDGEEDQEWIIGHIDGEPKRKGAPPVSVFHFAD"
```

ORIGIN

Alignment Scores:  
Pred. No.: 2,16e-153 Length: 4713  
Score: 161.00 Matches: 694

```
Percent Similarity: 95.99% Conservative: 0
Best Local Similarity: 95.99% Mismatches: 10
Query Match: 16.00% Indels: 29
DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x AKI72944 (1-4713)

Qy      1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAsePTrLysAla 20
Db      101 ATGCCGGACACAGATCTCCGTGTCGGAATTCGTGGCCGAGACCCATGAGGACTACAGCGC 160
Qy      21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db      161 CCCACGGCCTCTAGCTTACCACCGCGCACCGCCAGTCCGCGGAACACCGTGGCGCCATC 220
Qy      41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db      221 GAGGAGGCTTTGGACGTGGACCGGATGTTCTCTACAAAATGAAGAAATCTGTGAAGGCA 280
Qy      61 IleAsnSer-SerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuG1 80
Db      281 ATCAATAT-CTCTGGGCTGGCACACGTGGAGAACGAGAACAGTACACCCAGGCTCTGGA 339
Qy      80 uLysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuL1 100
Db      340 GAAGTTCGGTGGCAACTGTGTGTGCAGAGATGATCCAGACTTAGGAAGCGCGTTTCTGAA 399
Qy      100 sPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeGlnAsnMetAg 120
Db      400 GTTCTCTGTGTTACCAAGGAGTTGACGGCACTCTTCAAAAACCTGATTCAGAAACATGAA 459
Qy      120 nAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysG1 140
Db      460 CAACATCATCTCGTTCCTTGGACAGTTGCTGAAAGGAGATCTGAAGAGAGTAAAGG 519
Qy      140 YAspLeuLysLysProPheAspLysAlaTTrpLysAspTyrGluThrLysIleThrLysI1 160
Db      520 GGATCTGAAAAGGCCCTTTGATNAAGCTTGAAGGAGACTATGAAACAAAAATAACCAAAAT 579
Qy      160 eGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysIleSerG1 180
Db      580 AGAAAGAGAGAAAAAGGAGCATGCTCAAGCTCCAGCGGATGATTCGTACTGAAATTAAGCGG 639
Qy      180 YAlaGluIleAlaGluMetGluLysGluArgPheGlnLeuGlnMetCysG1 200
Db      640 GGCTGAGATTCCAGAGAGATGGAGAAAGAAAGACGGTTCTTCCAGCTGCAGATGTGTGA 699
Qy      200 uTyrLeuLeuLysValAsnGluIleLysIle-LysLysGlyValAspLeuLeuGlnAsnL 220
Db      700 GTACCTGCTGAAGGTCAATGAAATCAAGGT-GAAAAAGGGAGTGGATTTGCTTCAGAAATC 758
Qy      220 euLlLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluS 240
Db      759 TGATCAAGTACTTTTCAGCCCGCAGTGCAATTTTTTTCAGGATGGATTCAGAACAGTAGAAA 818
Qy      240 erLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnA 260
Db      819 GCCTCAAGCCTTCCATCGAGACGCTCTCCAGGACCTCCACACCATCAACAGGCCCCAGG 878
Qy      260 spGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValG 280
Db      879 ATGAGGAACCGCGACAGCTGATACAACATTCGAGATATTTTCAAGTCAGCATTCGAGGTGG 938
Qy      280 luGlnLysGlu-----AspSerGlnIle-ArgGlnSerThrAlaTyrSerLeuHis 296
Db      939 AACAGAAAGATCTAGGAGAGACTCACAACCT-TCGCCAAAGTACAGCCTCAGCTTACAT 997
Qy      297 GlnProGlnGlyAsnLysGluHisGlyThrGluArgAsn-GlySerLeuTyrLysLysSe 316
Db      998 CAACCTCAGGAAACAGGAGCATGGAAACAGAGAGAACCGGAAC-CTCTTACAGAGAG 1056
Qy      316 rAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuTh 336
```



Db 1057 CGATGGGATCCGAAAGTGTGGCAGAGAGGAAGTGTTCGCTTAAATAATGCTTCCTCCTAC 1116

Qy 336 rilesrhisglThrAlaAenArgProProAlaLysLeuAenLeuThrCysGlnVa 356

Db 1117 CATCTCCACGGCACTGCGCAACCGGCCACCTGCCAAGCTCAACCTGTCTAACTGCCAGGT 1176

Qy 356 llystranProGluGluLysLysCysPheAspLeuLysSerHisAspArgThrTyRhi 376

Db 1177 GAAGACCAATCCGAGGAGAGAAAGTGTTCGACCTCATATACATGATGACGAGCGTATCCA 1236

Qy 376 spheGlnAlaGluAspGluGlnGluCysGlnIleTrrMetSerValLeuGlnAenSerLys 396

Db 1237 CTTCAAGCGAGAGCAGCAGAGATGTCAGATATGGATGTCTGTACTGTCAGAACAGCA 1296

Qy 396 sGluGluAlaLeuAenAlaPheLysGlyAspAenThrGlyGluAenAenIleVa 416

Db 1297 GGAAGAAGCTCTGAACAACGCGCTTTAAGGGTGTGACAAACACTGGAGAAAATAACATCGT 1356

Qy 416 lGlnGluLeuThrLysGluLysLysLysSerGluValGlnArgMetThrGlyAenAspValCy 436

Db 1357 CCAAGAGCTGACCAAGGAGATCATCTCGAGAGGTGCAGAGGATGACGGCAATGACGTGTG 1416

Qy 436 sCysAspCysGlyAlaProAspProThrTrrLeuSerThrAenLeuGlyLysLeuThrCy 456

Db 1417 CTGGACTGTGGGGCACCAGATCCGACGTGCTCTACCAACCTGGGCATCTCGACTGTG 1476

Qy 456 sileGluCysSerGlyLysArgGluLeuGlyValHisTyrrSerPro-MetGlnSerL 476

Db 1477 CATCGAGTGTCTGGATCCACCGGGAGCTGGGGTTCATTACTCCAG-GATGAGTCCC 1535

Qy 476 euThrLeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAenIleGlyAenAlaG 496

Db 1536 TGACATTAGATGATTGGGAACGTCTGAGCTTCTGCTTCCCAAGAATATCGGAATGCGAG 1595

Qy 496 lPheAenGluLysMetGluCysCysLeuProAla-GluAspSer-ValLysProAspPr 515

Db 1596 GCTTTAATGAATATATGAGGTGTGCTTACCCTGTC-TGAGGACCC-AGTCAAGCCCCAACCC 1653

Qy 515 oGlySerAspMetAen-AlaArgLysAspTyrrIleThrAlaLysTyrrIle-GluArgArg 534

Db 1654 AGGCAGTGACATGAT-TCCAGAGGAGGACTACATCACAGCCAGTACAT-TGAGAGGAGA 1711

Qy 535 TyrAlaArgLysLysHisAlaAspAen-AlaAlaLysLeuHisSerLeuCysGluAlaVa 554

Db 1712 TACGCACGGAAGACATGACAGACAC-CGCGCGCAAGCTCCACAGCCCTTTCGAGGCCGT 1770

Qy 554 llystranProGluGluLysLysLysLysSerGluValAlaAspGlyValAspLeuThrGl 574

Db 1771 CAAGACCAAGACATTTTGGGTACTCCAAGCTTATCTGATGTGTGGACCTGACAGA 1830

Qy 574 uLysIleProLeuAlaAenGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr 594

Db 1831 GAAATCCCATGCGCCATGGCATGGCATGACAGATGAGACAGCCCTCCATCTTGAGTTCAG 1890

Qy 594 gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAenSerGlyAenLe 614

Db 1891 ATCTGTGACCGGACTTCCCTTCATTTAGACTTCTCGTCCAGAACAGTGGGAACCT 1950

Qy 614 uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrrCysCysLeuThrAspAenAl 634

Db 1951 GGATAAACAGACAGCAGGCGGAGCAGACAGCCCTGCTGCTGCTGACAGCAATGCG 2010

Qy 634 aGluCysLeuLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 654

Db 2011 TGAGTGCCTCAAGCTTCTCTCGGGGAGAAAGCCCTCCATTGATGATGATGATGATGATG 2070

Qy 654 yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuTh 674

Db 2071 AGAGACCCCGTTGGACATTTGCTAAGCGCTCAAGCAGCAACACTGTGAGGAACCTGCTCAC 2130

Qy 674 rGlnAlaLeuSerGlyArgPheAenSerHisValHisValGluTyrGluTrrArgLeuLe 694

Db 2131 TCAGGCCCTGTGCGGGCGGCTTTAACTCCACGTTCCAGTTCAGTATGATGATGATGATGATG 2190

Qy 694 uHisGluAspLeuAspGluSerAspAsp 704

Db 2191 GCATGAGGACCTGATGAGAGCGACGACGAT 2221

## RESULT 6

BC080847

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC080847 5072 bp mRNA linear ROD 20-SEP-2004  
Mus musculus gene model 592, (NCBI), mRNA (cdna clone MGC:90837  
IMAGE:6852788), complete cds.

BC080847 GI:51593636  
MGC.  
Mus musculus (house mouse)

1 (bases 1 to 5072)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalusz, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 5072)  
Director MGC Project.

Direct Submission  
Submitted (27-AUG-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mcg@nih.gov](mailto:nisc.mcg@nih.gov)

Akter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granito, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgoun, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, S.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: TRAK Plate: 176 Row: i Column: 16  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, GenomeScan gene prediction.

REMARK  
COMMENT



## ORIGIN

## Alignment Scores:

Pred. No.: 5.27e-122 Length: 2109  
 Score: 130.00 Matches: 130  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 12.92% Indels: 0  
 DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x CQ720712 (1-2109)

Qy 342 AlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGlu 361  
 Db 772 GCTAACCGGCTCTCTGCAAGCTCAACCTGCTAACTGCTGCAAGTGAAGACCAACCTGAG 831  
 Qy 362 GluLysLysCysPheAspLeuLysHisAspArgThrTyHisPheGlnAlaGluAsp 381  
 Db 832 GAGAGAAAGTCTTGGACCTTATTCATCAGCAGAACTTACCACCTTTCAAGCTGAAGAT 891  
 Qy 382 GluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsn 401  
 Db 892 GAACAGGAATGTCAATATGATGTCTGTGCTGCAAAATAGCAAGAGAGCTTTAAAC 951  
 Qy 402 AsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys 421  
 Db 952 AATGCAATTAAGGGGGATGACAATACCTGAGAAATAATACATCTGCTCAAGAACTGACAAAG 1011  
 Qy 422 GluLleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAla 441  
 Db 1012 GAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGTGATGTGGGGCG 1071  
 Qy 442 ProAspProThrTrpLeuSerThrAsnLeuGlyLleLeuThrCysLleGluCysSerGly 461  
 Db 1072 CCAGATCCTATGCTTTCACCAACCTGGGCATCTGACCTGCATCGAGTGTTCGGA 1131  
 Qy 462 IleHisArgGluLeuGlyValHisTySer 471  
 Db 1132 ATCCACCGAGAGCTGGGGGTTTCATTATTCC 1161

## RESULT 8

## AX746789

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## location/Qualifiers

## 1. 3123

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

## Pred. No.:

## Score:

## Percent Similarity:

## Best Local Similarity:

## Query Match:

## Length:

## Matches:

## Conservative:

## Mismatches:

## Indels:

## DB:

US-09-914-042-1 (1-1006) x AX746789 (1-3123)

Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860  
 Db 272 CCCTTGACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 331  
 Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880  
 Db 332 AGCCAGCGGAGCAAGCTGCCCCGCTGGGATCTCACAGATCAGGCCCGCCCGCCCGCC 391  
 Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSerTh 900  
 Db 392 CCACAGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 450  
 Qy 900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920  
 Db 451 CCCACTGACCAACAAGGCGCCCAAGGAGGACCTGTGGATCTCTCTGCAACGGAGCTCT 510  
 Qy 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940  
 Db 511 GGGTCCTCTGTCCCAATGCTATGTCCTGCAGCCCGCCCGCCCGCCCGCCCGCCCGCC 570  
 Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyTrpAsnCysValAlaAspAsnPr 960  
 Db 571 GGCAACCAAGTTGAAGCTTAAGCGGGTGAAGCGCTCTATTAATCTGTGTGGCTGACACCC 630  
 Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980  
 Db 631 CGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGAGCGGGAGGAGGACGAGA 690  
 Qy 980 uTrpTrpLleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
 Db 691 GTGGTGATTTGGCCACATTTGATGGAGATCTCTGGTCGCANAGCGCATCTCCGGTGTCT 750  
 Qy 1000 eValHisPheIleAlaAsp 1006  
 Db 751 TGTGCACTTTATCGCTGAC 769

## RESULT 9

## AK091121

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## location/Qualifiers

## 1. 3123

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

## Pred. No.:

## Score:

## Percent Similarity:

## Best Local Similarity:

## Query Match:

## Length:

## Matches:

## Conservative:

## Mismatches:

## Indels:

AX746789 3123 bp mRNA linear PAT 20-JUN-2003

Sequence 314 from Patent EP1308459.

AX746789

AX746789.1 GI:32131177

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,

Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,

Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and

Masuho, Y.

Full-length cDNA sequences

Patent: EP 1308459-A 314 07-MAY-2003;

Helix Research Institute (JP); Research Association for

Biotechnology (JP)

Location/Qualifiers

1. 3123

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

Length:

Matches:

Conservative:

Mismatches:

Indels:

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetake, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizuehima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)

14702039

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T., Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R., Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, E., Momiyama, H., Onogawa, S., Kaeriyama, S., Satoh, N., Matsunawa, H., Takahashi, E., Kataoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S., Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Satoh, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished

3 (bases 1 to 3123)

Isogai, T. and Yamamoto, J.

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan, cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES  
source

Location/Qualifiers  
1..3123  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CTONG2000452"  
/tissue\_type="tongue, tumor tissue"  
/clone\_lib="CTONG2"  
/note="Cloning vector: pME18SFL3"  
146..772  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAC03588.1"  
/db\_xref="GI:21749416"

CDS

/translation="MPSTLQNETYCALLSGSPPPAQAAPASTTAPPLPPRNVGKPL  
TPTTPPVAKTFSVMSALQSPKAPGISOIRPPPLPPQPSRLPQKPPAPGADKST  
PLTNKGQPRGVDLSATLGLSNMVLQPPAPMPKRSQATKLKPKRVKALYNVCVAD  
NPDLTSEGDVIIVDGEDQEWIGHIDGDPGRKGAPFVSFVFIAD"

ORIGIN

Alignment Scores:  
Pred. No.: 2,026-101 Length: 3123  
Score: 110.00 Matches: 165  
Percent Similarity: 98.80% Conservative: 0  
Best Local Similarity: 98.80% Mismatches: 1  
Query Match: 10.93% Indels: 2  
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x AK091121 (1-3123)

Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
Dbb 272 CCCCTGACCCCGCCGCGCCCGCCCGCTTCCCAAGACGCCCGCGTAATGGAAGCCTTG 331  
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnLysArgProProProLeuPro 880  
Dbb 332 AGCCAGCGGAGCAAGCTGCTCCCGCCCTGGGATCTCAGACATCAGCGCCCGCCACCTCTGCC 391  
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProProGlyThrLysSerTh 900  
Dbb 392 CCACAGCGCGCCGCGCTCCCGCAGAGAAGCTCGCGCGGGGCT-GACAAAGTCCAC 450  
Qy 900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920  
Dbb 451 CCCACTGACCAACAAAGCCCAACCGAGAGGACCTGTGGATCTCTCTCAACGGAAGCTCT 510  
Qy 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGl 940  
Dbb 511 GGGTCCTCTGTCNAATGCTATGTCCTGCAGCCCGCTGCACCATGCTAGGAAGTCGCA 570  
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysValAlaAspAsnPr 960  
Dbb 571 GGCACCAAGTTGAAGCTTAAGCGGGTGAAGCGCTCTATAACTGTGTGGCTGACAAACC 630  
Qy 960 oAspGluLeuThrPheSerGluGlyAspValLleValAspGlyGluGluAspGlnGl 980  
Dbb 631 CGATGAGCTCACCTTCTCCGAGGGGGATGTATCATCTGTGACCGGGAGGACACAGGA 690  
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
Dbb 691 GTGGTGGATTGGCCACATTCATGAGAGATCTCTGGTGCAGAGGCGCATTCCTCGGTGTCTATT 750  
Qy 1000 eValHisPheIleAlaAsp 1006  
Dbb 751 TGTGCACTTTATGCTGCTGAC 769  
RESULT 10  
AK124900  
LOCUS  
DEFINITION Homo sapiens cDNA FLJ42910 fis, clone BRHIP3021778, highly similar to Homo sapiens development and differentiation enhancing factor 2 (DDEF2).  
ACCESSION AK124900  
VERSION AK124900.1 GI:34530818  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayaashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Iehii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yanazaki, M., Ninomiya, K., Ishibashi, T., Yanashita, H., Muraoka, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirao, K., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togli, S., Komai, F., Hara, R., Harada, H., Tanase, T., Imose, N., Mutsashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satchi, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuko, Y., Yamaashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039

# TITLE JOURNAL PUBMED REFERENCE AUTHORS

Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
3 (bases 1 to 3870)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUN-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

# TITLE JOURNAL PUBMED REFERENCE AUTHORS

Location/Qualifiers  
1. 3870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="BRHIP3021778"  
/tissue\_type="hippocampus"  
/clone\_lib="BRHIP3"  
/note="cloning vector: pME18SFL3"

# FEATURES source

Alignment Scores:  
Pred. No.: 2.42e-101 Length: 3870  
Score: 110.00 Matches: 110  
Percent Similarity: 100.00% Conservatative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 10.93% Indels: 0  
DB: 9 Gaps: 0

# ORIGIN

US-09-914-042-1 (1-1006) x AK124900 (1-3870)  
QY 897 AsplVsSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAla 916  
DB 1187 GACAGTCCACCCCACTACCAACAAAGCCACCGAGAGGACCTGTGGATCTCTCTCA 1246  
QY 917 ThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetPro 936  
DB 1247 ACGGAAGCTCTGGGTCTCTGTGCAATGCTATGTCCTGCAGCCCTGCACCATGCT 1306  
QY 937 ArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuThrAsnCysVal 956  
DB 1307 AGGAAGTCCGAGGCAACCAAGTTGAAGCTTAAGCGGGTGAAGCGCTCTATACTGTG 1366  
QY 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleileValAspGlyGlu 976  
DB 1367 GCTGACACCCCGATAGACTACCTTCCTCCGAGGGGATGATCATCTGTGACGGGAG 1426  
QY 977 GluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996

DB 1427 GAGGACCAAGAGTGGTGGATTGGCCACATTGATGAGATCTGTCGAAAGCGCATTC 1486  
QY 997 ProValSerPheValHisPheIleAlaAsp 1006  
DB 1487 CCGGTGTCATTGTGCACTTTATCGCTGAC 1516

# RESULT 11

CR385625 LOCUS CR385625 858 bp mRNA linear VRT 05-APR-2004  
DEFINITION Gallus gallus finished cDNA, clone CHEST386c18.  
ACCESSION CR385625  
VERSION CR385625.1 GI:46238384  
KEYWORDS  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

# REFERENCE

1 (bases 1 to 858)  
AUTHORS Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E., Croning, M.D.R., Davies, R.M., Francis, M.D., Grafham, D.V., Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R., Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G., Tickle, C. and Wilson, S.A.  
DIRECT SUBMISSION  
Submitted (05-APR-2004) Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: chicken@bma.umist.ac.uk  
COMMENT BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA sequencing project.  
This sequence is from the BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection, from a library constructed by Elizabeth Bosch. cDNA was prepared from RNA extracted from heads, normalised, and poly A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector. Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia coli DH10B.

# TITLE

JOURNAL  
COMMENT  
BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA sequencing project.

# FEATURES

Location/Qualifiers  
1. 858  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hiseex"  
/db\_xref="taxon:9031"  
/clone="CHEST386c18"  
/clone\_lib="CSEQCHN62"  
/dev\_stage="stage 36"

# ORIGIN

Alignment Scores:  
Pred. No.: 7.55e-100 Length: 858  
Score: 108.00 Matches: 160  
Percent Similarity: 98.77% Conservatative: 0  
Best Local Similarity: 98.77% Mismatches: 1  
Query Match: 10.74% Indels: 2  
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x CR385625 (1-858)

QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
DB 345 CAGAAAGAGGATTCAGATTTCGTGAGATTCAGTACAGTTTACATTCAGCTCAGGGA 404  
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
DB 405 AACAGGACATGCATGAACGAATGGAGTCTGTACAAGAAAGTGAATGGAATCAGA 464  
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
DB 465 AAAGTGTGGCAGAAAAGGAGTCTAGTAAAAACGGTTTTCTTACAAATTTCCCATGGT 524  
QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360  
DB 525 ACAGCTAACCGACCTCCAGCAAGCTCAATCTGTTAACTCTGCAAGTGAACCAACCA 584

```

Qy 361 GluGluYsLYsCysPheAspLeuIleSerHisApArgThrTyTyrHisPheGlnAlaGlu 380
Db 585 GAGGAGAAAATAATGTTTGGCTCATATCATGACAGGACATACCACTTCCAAGCAGAA 644
Qy 381 AspGluGlnGluCysGlnIleTrpMet-SerValLeuGlnAenSerLYsGluGluAlaLe 400
Db 645 GATGACAGGAAATGTCATATATGGAC-ATCTGTTCTACAAAACACAGAGGAGGAACTTT 703
Qy 400 uAsnAenAlaPheLYsGlyAspAspAsnThrGlyGluAenAenIleValGlnGluLeuTh 420
Db 704 AAATAATGCATTCAAAGGAGATGATACACAGGAGAAAAATAATATTGTCAGGAACGAC 763
Qy 420 rLYsGluIleIleSerGluValGlnArgMetThrGlyAenAspValCysCysAspCysG1 440
Db 764 AAAAGAGATTATATCTGAAGTCCAAAGGATGACTGGAATGATGTGTGTGACTGTGG 823
Qy 440 YAla 441
Db 824 AGCA 827

RESULT 12
AC080162
LOCUS
DEFINITION Homo sapiens BAC clone RP11-400L8 from 2, complete sequence.
ACCESSION AC080162
VERSION AC080162.7 GI:18093140
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE
2 (bases 1 to 171653)
Cordes,M., Kozlowski,A., Dixon,R. and Boyer,E.
The sequence of Homo sapiens BAC clone RP11-400L8
Unpublished (2001)
3 (bases 1 to 171653)
Waterston,R.H.
Direct Submission
Submitted (27-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 171653)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:15145269.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0400L08
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-385J23; the clone sequenced to the right is RP11-264H21. Actual start of this clone is at base position 1 of RP11-400L8; actual end is at base position 171653 of RP11-400L8.

Data from AC079782 was used to finish this clone, AC080162.

Sequence derived from a single plasmid subclone from base position 47301 to 47329.

FEATURES	source	Location/Qualifiers
	1. 171653	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone="RP11-400L8"
		/clone_lib="RPCI-11"
misc_feature	179..418	
repeat_region	346..627	/note="match to EST BG959460 (NID:gl14377631)"
	741..1147	/rpt_family="Alu"
misc_feature	890..1053	/note="match to EST BF929212 (NID:gl2327340)"
repeat_region	1539..1734	/rpt_family="Alu"
	1707..1734	/rpt_family="Alu"
repeat_region	2030..2054	/rpt_family="(A)n"
repeat_region	2079..2490	/note="family=AT-rich"
	3105..3156	/rpt_family="L1"
misc_feature	3115..3156	/note="similar to Sus scrofa EST BG384893 (NID:gl3309365)"
misc_feature	3120..3156	/note="match to EST C16709 (NID:gl571416)"
misc_feature	3240..3406	/note="similar to Rattus norvegicus EST BF394200 (NID:gl1379064)"
misc_feature	3240..3406	/note="match to EST C16709 (NID:gl571416)"
misc_feature	3240..3406	/note="similar to Rattus norvegicus EST BF394200 (NID:gl1379064)"
misc_feature	3240..3406	/note="similar to Sus scrofa EST BG384893 (NID:gl3309365)"
misc_feature	3312..3561	/note="match to EST BF751012 (NID:gl2077688)"
misc_feature	3384..3406	/note="similar to Homo sapiens EST BF804201 (NID:gl2133190)"

```

misc_feature      3731..4236
/note="match to EST BF055560 (NID:g10809456)"
repeat_region    3805..3897
/rpt_family="MIR"
misc_feature      4004..4240
/note="match to EST AI708030 (NID:g4997806) as60b02.xl"
repeat_region    4258..4557
/rpt_family="Alu"
misc_feature      5817..5956
/note="similar to Rattus norvegicus EST BF394200
(NID:g11379064)"
misc_feature      5817..5950
/note="match to EST AW996731 (NID:g8256965)"
misc_feature      5817..5950
/note="similar to Homo sapiens EST BF804201
(NID:g12133190)"
misc_feature      5817..5950
/note="similar to Sus scrofa EST BG384893 (NID:g13309365)"
repeat_region    5817..5900
/note="match to EST C16709 (NID:g1571416)"
misc_feature      7266..7491
/rpt_family="MIR"
misc_feature      7385..7734
/note="match to EST AA377813 (NID:g2030152)"
repeat_region    7596..7978
/rpt_family="MER1_type"
repeat_region    8578..8648
/rpt_family="T-rich"
repeat_region    8620..8924
/rpt_family="Alu"
misc_feature      8905
/note="match to EST H70683 (NID:g1042499) yul9f04.r1"
misc_feature      8942..9396
/note="match to EST H70683 (NID:g1042499) yul9f04.r1"
repeat_region    9033..9073
/rpt_family="(TG)n"
repeat_region    9454..9481
/rpt_family="AT-rich"
repeat_region    9479..9617
/rpt_family="L1"
repeat_region    9618..9911
/rpt_family="Alu"
repeat_region    9912..10645
/rpt_family="L1"
repeat_region    10646..10924
/rpt_family="Alu"
repeat_region    11287..11577
/rpt_family="Alu"
repeat_region    11870..12043
/rpt_family="MIR"
repeat_region    14085..14106
/rpt_family="AT-rich"
repeat_region    14089..14392
/rpt_family="Alu"
repeat_region    14540..14833
/rpt_family="Alu"
repeat_region    14936..15102
/rpt_family="MIR"
misc_feature      15446..15580
/note="match to EST H56609 (NID:g1005253) yc88d10.r1"
misc_feature      15486..15580
/note="match to EST AW996731 (NID:g8256965)"
misc_feature      15486..15580
/note="similar to Sus scrofa EST BG384893 (NID:g13309365)"
misc_feature      15486..15566
/note="similar to Homo sapiens EST BF804201
(NID:g12133190)"
misc_feature      16075..16092
/note="match to EST AI831610 (NID:g5452281) wj39g03.xl"
repeat_region    16079..16378
/rpt_family="Alu"
repeat_region    16937..17197
/rpt_family="Alu"

```

```

misc_feature      17082..17601
/note="match to EST BF829427 (NID:g12174985)"
repeat_region    17585..17753
/rpt_family="MIR"

Alignment Scores:
Pred. NO.:      2.53e-70      Length:      171653
Score:          81.00      Matches:      81
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     8.05%      Indels:      0
DB:              9      Gaps:      0

US-09-914-042-1 (1-1006) x AC080162 (1-171653)

Qy 714 AsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSer 733
|||||
Db 35364 AACCGGGGGGAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCACCACGCTTCAGTCT 35423

Qy 734 AsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPhe 753
|||||
Db 35424 AACGCTGTATCTTTGGCCAGAGATGCTGCAACCTTGCAAGGAGAGACGAGGGCTTTC 35483

Qy 754 MetProSerIleLeuGlnAsnGluThrTyrClyAlaLeuLeuSerGlySerProProPro 773
|||||
Db 35484 ATGCCCGAGCATCTTGCAGAAATGAGACTTACGAGCCCTCTCTGAGTGGCAGCCACCTCCC 35543

Qy 774 AlaGlnProAlaAlaProSerThrSerAlaProProLeuProProAlaAsnValGly 793
|||||
Db 35544 GCCCAGGCTGCAGCCGCCAGCACCACCGGCCCGCTTCTCCACGGAATGTTGGC 35603

Qy 794 Lys 794
|||||
Db 35604 AAA 35606

RESULT 13
AKI23185
LOCUS
DEFINITION Homo sapiens cDNA FLJ41191 fis, clone BRACE2044946, weakly similar
to Homo sapiens ADP-ribosylation factor (arf)-directed GTPase
activating protein.
ACCESSION AKI23185
VERSION AKI23185.1 GI:34528666
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hottuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aota, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Tarashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

```

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oiyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039  
2

**TITLE**  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Waga, S., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2157)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
**FEATURES**  
source  
1. .2157  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="BRACE2044946"  
/tissue\_type="cerebellum"  
/clone\_lib="BRACE2"  
/note="cloning vector: pME18SFL3"

**ORIGIN**

Alignment Scores:  
Pred. No.: 1.07e-61 Length: 2157  
Score: 71.00 Matches: 71  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.06% Indels: 0  
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x AK123185 (1-2157)

Qy 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTyrMetSerVal 391  
|  
|  
|  
Db 1113 GACGAACTTACCCTTCAGCTGAAGATGAACAGGAATGTCAAATATGATGTCGTG 1172  
|  
|  
|

Qy 392 LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411  
|  
|  
|  
Db 1173 CTGCAAAATAGCAAGAGAGAGCTTTAAACAATGCAATTAAGGGGGATGACAATACTGGA 1232  
|  
|  
|

Qy 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431  
|  
|  
|  
Db 1233 GAAATAACATCGTCCAGAACTGACAAAGGAGATCATCTCAGAGTGCAGAGATGACG 1292  
|  
|  
|

Qy 432 GlyAsnAspValCysCysAspCysGlyAlaPro 442  
|  
|  
|  
Db 1293 GGCAATGACCTGCTGTGACTGTGGGGCGCCA 1325  
|  
|  
|

RESULT 14  
AR243296  
LOCUS

2949 bp DNA linear PAT 20-DEC-2002

DEFINITION Sequence 8 from patent US 6475778.  
ACCESSION AR243296  
VERSION AR243296.1 GI:27290454  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2949)  
AUTHORS Roberts, T.M., King, F.J., Harris, D.F., Hu, E., Spiegelman, B. and Chan, J.  
TITLE Differentiation enhancing factors and uses therefor  
JOURNAL Patent: US 6475778-A 8 05-NOV-2002;  
FEATURES Location/Qualifiers  
source 1. .2949  
/organism="unknown"  
/mol\_type="genomic DNA"

**ORIGIN**

Alignment Scores:  
Pred. No.: 1.61e-58 Length: 2949  
Score: 68.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.76% Indels: 0  
DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x AR243296 (1-2949)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyValaGluIleAlaGluMetGluLysGlu 190  
|  
|  
|  
Db 511 CACGGAATGATCCGACGGAGATCAGCGAGATAGCAGAGATAGCAGAGATGGAAGAGAG 570  
|  
|  
|

Qy 191 ArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLysValAsnGluIleLysIle 210  
|  
|  
|  
Db 571 CGCGCTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAGTCAATGAATCAAGATC 630  
|  
|  
|

Qy 211 LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe 230  
|  
|  
|  
Db 631 AAAAAAGGTGCGACCTGCTCCAGAACTCATCAATCTCCACGACAGTGCAACTTC 690  
|  
|  
|

Qy 231 PheGlnAspGlyLeuLysAlaVal 238  
|  
|  
|  
Db 691 TTTCCAGATGTCTCTCAAGCGGTG 714  
|  
|  
|

RESULT 15  
AR243295  
LOCUS

DEFINITION Sequence 6 from patent US 6475778.  
ACCESSION AR243295  
VERSION AR243295.1 GI:27290453  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5954)  
AUTHORS Roberts, T.M., King, F.J., Harris, D.F., Hu, E., Spiegelman, B. and Chan, J.  
TITLE Differentiation enhancing factors and uses therefor  
JOURNAL Patent: US 6475778-A 6 05-NOV-2002;  
FEATURES Location/Qualifiers  
source 1. .5954  
/organism="unknown"  
/mol\_type="genomic DNA"

**ORIGIN**

Alignment Scores:  
Pred. No.: 2.91e-58 Length: 5954  
Score: 68.00 Matches: 68  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.76% Indels: 0  
DB: 6 Gaps: 0



US-09-914-042-1 (1-1006) x AR243295 (1-5954)

Qy	171	HisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGlu	190
Db	943	CACGGAATGATCCGACGAGATCAGCGGACGAGATAGCAGAAGAGATGGAAGAG	1002
Qy	191	ArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIle	210
Db	1003	CGGCGTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAAGTCAATGAATCAAGATC	1062
Qy	211	LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe	230
Db	1063	AAAAAGGTGTCGACCTGCTCCAGAACTCATCAAAATCTTCCACGACAGTGCAACTTC	1122
Qy	231	PheGlnAspGlyLeuLysAlaVal	238
Db	1123	TTTCAGGATGGTCTCAAGCGGTG	1146

Search completed: August 4, 2005, 22:16:00  
Job time : 10766 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 00:25:21 ; Search time 113 Seconds  
(without alignments)  
4558.864 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 1006  
Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAFPVSFVHFIAID 1006

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt\_03.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	703	69.9	1006	1	DDF2_HUMAN	O43150 homo sapien
2	161	16.0	784	1	DDF2_MOUSE	Q78196 mus musculus
3	161	16.0	809	2	Q66JN2	Q66JN2 mus musculus
4	161	16.0	970	2	Q6A074	Q6A074 mus musculus
5	110	10.9	208	2	Q8N282	Q8N282 homo sapien
6	46	4.6	1029	2	Q6NRK0	Q6NRK0 xenopus lae
7	34	3.4	108	2	Q8K115	Q8K115 mus musculus
8	30	3.0	1129	1	DDF1_BOVIN	O97902 b 130-kda p
9	30	3.0	1147	1	DDF1_MOUSE	Q9Gwy8 m 130-kda p
10	28	2.8	111	2	Q8BYE8	Q8BYE8 mus musculus
11	23	2.3	98	2	Q8K0U0	Q8K0U0 mus musculus
12	22	2.2	956	1	DDF1_HUMAN	Q9ulh1 h 130-kda p
13	19	1.9	123	2	Q8C1G6	Q8C1G6 mus musculus
14	16	1.6	121	2	Q8R1L5	Q8R1L5 mus musculus
15	15	1.5	155	2	Q9D308	Q9D308 mus musculus
16	14	1.4	104	2	Q9CYR1	Q9CYR1 mus musculus
17	14	1.4	112	2	Q8P562	Q8P562 mus musculus
18	14	1.4	192	2	Q8C3B6	Q8C3B6 mus musculus
19	12	1.2	1243	2	Q8TBN3	Q8TBN3 homo sapien
20	12	1.2	1243	2	Q6T7X3	Q6T7X3 homo sapien
21	12	1.2	1244	2	O00562	O00562 homo sapien
22	12	1.2	1244	2	Q9BZ73	Q9BZ73 homo sapien
23	11	1.1	307	2	Q8DEH1	Q8DEH1 brachydanio
24	10	1.0	118	2	Q8C0H9	Q8C0H9 mus musculus
25	10	1.0	138	1	YGJM_ECOL6	P67702 escherichia
26	10	1.0	138	1	YGJM_ECOL1	P67701 escherichia
27	10	1.0	138	1	YGJM_SHIFL	P67703 shigella fl
28	10	1.0	903	2	Q8TDY4	Q8TDY4 homo sapien
29	10	1.0	903	2	Q8P9F4	Q8P9F4 homo sapien
30	10	1.0	1358	2	Q7RWZ2	Q7RWZ2 neurospora
31	10	1.0	1396	2	Q8WZY9	Q8WZY9 neurospora

32 9 0.9 125 2 Q8H4F8  
33 9 0.9 202 2 Q6PJ46  
34 9 0.9 281 2 Q95UG6  
35 9 0.9 287 2 Q95UG5  
36 9 0.9 309 2 Q9P6R1  
37 9 0.9 322 2 Q8K0K1  
38 9 0.9 356 2 Q9H3X2  
39 9 0.9 360 2 Q96T14  
40 9 0.9 366 2 Q811F3  
41 9 0.9 403 2 Q86XV5  
42 9 0.9 487 2 Q6GNT5  
43 9 0.9 513 2 Q7RHW9  
44 9 0.9 519 2 Q6PA21  
45 9 0.9 529 2 Q8BV81

#### ALIGNMENTS

RESULT 1  
ID DDF2\_HUMAN STANDARD; PRT; 1006 AA.  
AC O43150;  
DT 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Development and differentiation-enhancing factor 2 (Pyk2 C-terminus  
DE associated protein) (PAP) (Paxillin-associated protein with ARFGAP  
DE activity 3) (PAG3).  
GN Name=DDEF2; Synonym=KIAA0400;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=98116655; PubMed=9455477;  
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VIII.  
RT 78 new cDNA clones from brain which code for large proteins in  
RT vitro.";  
RL DNA Res. 4:307-313 (1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Smit  
FAhey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Trimmick J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinaki M.I., Skalak U., Smalls D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP FUNCTION, PHOSPHORYLATION, INTERACTION WITH ARF1; ARF5; ARF6; PTK2B  
RP AND SRC, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE  
RP SPLICING.  
RX PubMed=10022920;

Q8H4F8 oryza sativ  
Q6PJ46 homo sapien  
Q95UG6 babesia bov  
Q95UG5 babesia bov  
Q9P6R1 schizosacch  
Q8K0K1 mus musculus  
Q9H3X2 homo sapien  
Q96T14 homo sapien  
Q811F3 mus musculus  
Q86XV5 homo sapien  
Q6GNT5 xenopus lae  
Q7RHW9 plasmodium  
Q6PA21 xenopus lae  
Q8BV81 mus musculus

RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,  
 RA Randazzo P.A., Schlessinger J.;  
 RT "Identification of a new Pyk2 target protein with Arf-GAP activity.";  
 RL Mol. Cell. Biol. 19:2338-2350(1999).  
 [4]  
 RP FUNCTION, MUTAGENESIS OF CYS-436, SUBCELLULAR LOCATION, AND  
 RP INTERACTION WITH PXN.  
 RX PubMed=1074932;  
 RA Kondo A., Hashimoto S., Yano H., Nagayama K., Mazaki Y., Sabe H.;  
 RT "A new paxillin-binding protein, PAG3/Papalpa/KIAA0400, bearing an  
 RT ADP-ribosylation factor GTPase-activating protein activity, is  
 RT involved in paxillin recruitment to focal adhesions and cell  
 RT migration.";  
 RL Mol. Biol. Cell 11:1315-1327(2000).  
 [5]  
 RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN  
 RP FILAMENTS.  
 RX PubMed=11304556;  
 RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;  
 RT "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-  
 RT ribosylation factor (ARF), regulates ARF6 in Fc gamma receptor-mediated  
 RT phagocytosis of macrophages.";  
 RL J. Exp. Med. 193:955-966(2001).  
 CC -!- FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6.  
 CC Regulates the formation of post-Golgi vesicles and modulates  
 CC constitutive secretion. Modulates phagocytosis mediated by Fc  
 CC gamma receptor and ARF6. Modulates FXN recruitment to focal  
 CC contacts and cell migration.  
 CC -!- SUBUNIT: Binds PXN, ARF1, ARF5, ARF6, PTK2B and SRC.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein  
 CC associated with the plasma membrane and with Golgi stacks.  
 CC Colocalizes with F-actin and ARF6 in phagocytic cups.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=PAPalpha;  
 CC IsoId=043150-1; Sequence=Displayed;  
 CC Name=2; Synonyms=PAPbeta;  
 CC IsoId=043150-2; Sequence=VSP\_009722;  
 CC -!- TISSUE SPECIFICITY: Detected in heart, brain, placenta, kidney,  
 CC monocytes and pancreas.  
 CC -!- INDUCTION: Up-regulated during monocyte maturation.  
 CC -!- DOMAIN: The conserved Arg-464 in the Arf-GAP domain probably  
 CC becomes part of the active site of bound small GTPases and is  
 CC necessary for Gnp hydrolysis.  
 CC -!- PTM: Phosphorylated on tyrosine residues by SRC and PTK2B.  
 CC -!- SIMILARITY: Contains 2 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 Arf-GAP domain.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB007860; BAA23696.2; ALT\_INIT.  
 DR EMBL; BC063308; AAH63308.1; -.  
 DR HSPG; O60631; IGBO.  
 DR Genew; HGNC:2721; DDEF2.  
 DR MIM; 603817; -.  
 DR GO; GO:0008047; F:enzyme activator activity; TAS.  
 DR ProDom; P0000066; SH3; 1.  
 DR PROSITE; PS50297; ANK REP REGION; 1.  
 DR PROSITE; PS50088; ANK REPEAT; 1.  
 DR PROSITE; PS50115; ARFGAP; 1.  
 DR PROSITE; PS50003; PH DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Alternative splicing; ANK repeat; Coiled coil; Golgi stack;  
 KW GTPase activation; Metal-binding; Phosphorylation; Repeat; SH3 domain;  
 KW Zinc.

FT	DOMAIN	305	397	PH.
FT	DOMAIN	421	543	Arf-GAP.
FT	REPEAT	584	616	ANK 1.
FT	REPEAT	620	649	ANK 2.
FT	DOMAIN	944	1006	SH3.
FT	DOMAIN	256	283	Coiled coil (Potential).
FT	DOMAIN	729	752	Coiled coil (Potential).
FT	DOMAIN	771	936	Pro-rich.
FT	VARSPLIC	795	840	VQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPLPLR
FT				VTSTN -> D (in isoform 2).
FT	MUTAGEN	436	436	/FTId=VSP_009722.
FT	CONFLICT	86	86	C -> R (in Ref. 2).
FT	CONFLICT	748	748	E -> D (in Ref. 2).
FT	SEQUENCE	1006 AA;	111650 MW;	6A213517DCD995E1B CRC64;
Query Match		69.9%;	Score 703;	DB 1; Length 1006;
Best Local Similarity		99.7%;	pred. No. 0;	
Matches 1003;	Conservative	0;	Mismatches	3; Indels
				0; Gaps
0;				
Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATEEALDVRMVLKMKSKVKA	60	
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATEEALDVRMVLKMKSKVKA	60	
Qy	61	INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN	120	
Db	61	INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN	120	
Qy	121	NIISFPLDSLKGLKGVKGLKPKPKAMKDYETKITKEKEKEHAKLHGMIRTEISG	180	
Db	121	NIISFPLDSLKGLKGVKGLKPKPKAMKDYETKITKEKEKEHAKLHGMIRTEISG	180	
Qy	181	AEIAEEMEKERRFQLOMCEYLLKVKNEIKKGVLDLQNLKIYFHAQCNFPQDGLKAVES	240	
Db	181	AEIAEEMEKERRFQLOMCEYLLKVKNEIKKGVLDLQNLKIYFHAQCNFPQDGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERQQLQRLDILKSAQVEOKESQIRQSTAYSLSHQPG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERQQLQRLDILKSAQVEOKESQIRQSTAYSLSHQPG	300	
Qy	301	NKEHGTNRNGSLYKKSIGIRKVKQKCSVKNQGLTISHGTANRPPAKNLNLTQVKTNP	360	
Db	301	NKEHGTNRNGSLYKKSIGIRKVKQKCSVKNQGLTISHGTANRPPAKNLNLTQVKTNP	360	
Qy	361	BEKCFDLISHDRTHYFQAEDQEQCIWMSVLQNSKEALNNAFKGDDNTGNNIVQELT	420	
Db	361	BEKCFDLISHDRTHYFQAEDQEQCIWMSVLQNSKEALNNAFKGDDNTGNNIVQELT	420	
Qy	421	KEIISVQRMGTGNDVCCGAPDPTWLSTNLGILTCIECSGIRHRELGVHVSQMOSLTLDV	480	
Db	421	KEIISVQRMGTGNDVCCGAPDPTWLSTNLGILTCIECSGIRHRELGVHVSQMOSLTLDV	480	
Qy	481	LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Db	481	LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Qy	541	ADNAKHLSCFAVKTEDI FGLLOAYADGVDTLTKIPLANGHEDETHALHLAVSRVDRTS	600	
Db	541	ADNAKHLSCFAVKTEDI FGLLOAYADGVDTLTKIPLANGHEDETHALHLAVSRVDRTS	600	
Qy	601	LHIVDFLVQNSGNDLKOTGKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETPLDI	660	
Db	601	LHIVDFLVQNSGNDLKOTGKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETPLDI	660	
Qy	661	AKRLKHEHCELLTQALSGRPNSHVVEYEWRLHEDLDESDDMDKELQSENNRDRP	720	
Db	661	AKRLKHEHCELLTQALSGRPNSHVVEYEWRLHEDLDESDDMDKELQSENNRDRP	720	
Qy	721	ISFYQLGSNQLQSNVSLARDAAANLAKKORAFMPSILQNETYCALLSGSPPPQAPAPS	780	
Db	721	ISFYQLGSNQLQSNVSLARDAAANLAKKORAFMPSILQNETYCALLSGSPPPQAPAPS	780	
Qy	781	TTSAPPLPRNVGVQVQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPLPLRVTSTN	840	

```

Db 781 TTSAPPLPRNVGKQVOTASSANTLWKTNSVVDGSSRSSDDPAVHPPLPPPLRVSTN 840
Qy 841 PLTPPPPPVAKTPSVMEALSQSPKAPPGISQIRPPPLPQPSPRLPQKPKAPGTDKST 900
Db 841 PLTPPPPPVAKTPSVMEALSQSPKAPPGISQIRPPPLPQPSPRLPQKPKAPGADKST 900
Qy 901 PLTNKGQPRGVDSLSATGALGPLSNAMVLQPPAPMPKRSQATKLKPKRVKALYNCAVADNP 960
Db 901 PLTNKGQPRGVDSLSATGALGPLSNAMVLQPPAPMPKRSQATKLKPKRVKALYNCAVADNP 960
Qy 961 DELTFSEGDVIVDGEEDQEWIGHIDGPKRGKAPFVSFVHFAD 1006
Db 961 DELTFSEGDVIVDGEEDQEWIGHIDGPKRGKAPFVSFVHFAD 1006

RESULT 2
DDF2_MOUSE STANDARD; PRT; 784 AA.
AC Q7SIG6;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Development and differentiation-enhancing factor 2 (Pyl2 C-terminus
DE associated protein) (PAP) (Paxillin-associated protein with ARFGAP
DE activity 3) (PAG3).
GN Name=Ddef2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP CLONING, FUNCTION, PHOSPHORYLATION, INTERACTION WITH PTK2B AND SRC,
RP AND SUBCELLULAR LOCATION.
RX PubMed=10022920;
RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,
RA Randazzo P.A., Schlessinger J.;
RT "Identification of a new Pyl2 target protein with Arf-GAP activity.";
RL Mol. Cell. Biol. 19:2338-2350(1999).
RN [2]
RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN
RP FILAMENTS
RP PubMed=11304556;
RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;
RA "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-
RT ribosylation factor (ARF), regulates ARF6 in Pcgamma receptor-mediated
RT phagocytosis of macrophages.";
RL J. Exp. Med. 193:955-966(2001).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 245-522 IN COMPLEX WITH A
RP ZINC ION, AND MUTAGENESIS OF TRP-274; ILE-285; ARG-292; LEU-306 AND
RP ASP-307.
RX MEDLINE=20069319; PubMed=10601011; DOI=10.1093/emboj/18.24.6890;
RA Mandiyan V., Andreev J., Schlessinger J., Hubbard S.R.;
RT "Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-
RT associated protein beta.";
RL EMBO J. 18:6890-6898(1999).
CC -!- FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6.
CC Regulates the formation of post-Golgi vesicles and modulates
CC constitutive secretion. Modulates phagocytosis mediated by Fc
CC gamma receptor and ARF6. Modulates PAXN recruitment to focal
CC contacts and cell migration (By similarity).
CC -!- SUBUNIT: Binds PAXN, ARF1, ARF5, ARF6, PTK2B and SRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with the plasma membrane and with Golgi stacks.
CC Colocalizes with F-actin and ARF6 in phagocytic cups.
CC -!- DOMAIN: The conserved Arg-292 in the Arf-GAP domain probably
CC becomes part of the active site of bound small GTPases and is
CC necessary for GTP hydrolysis.
CC -!- PTM: Phosphorylated on tyrosine residues by SRC and PTK2B (By
CC similarity).
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.

```

```

CC -!- SIMILARITY: Contains 1 PH domain.
CC PDB; IDQ; X-ray; A=1-278.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGAP; 1.
DR PRODOM; PD000066; SH3; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50088; ANK REPEAT; 1.
DR PROSITE; PS50115; ArfGAP; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW 3D-structure; ANK repeat; Coiled coil; Golgi stack; GTPase activation;
KW Metal-binding; Phosphorylation; Repeat; SH3 domain; Zinc.
FT DOMAIN 84 111
FT DOMAIN 133 225
FT DOMAIN 249 371
FT REPEAT 412 444
FT REPEAT 448 477
FT DOMAIN 599 714
FT DOMAIN 722 784
FT MUTAGEN 274 274
FT MUTAGEN 285 285
FT MUTAGEN 292 292
FT MUTAGEN 292 292
FT MUTAGEN 306 306
FT MUTAGEN 307 307
FT SEQUENCE 784 AA; 87050 MW; F6775CFD7870353 CRC64;
Query Match 16.0%; Score 161; DB 1; Length 784;
Best Local Similarity 100.0%; Pred.No. 4e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 603
Db 372 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 431
Qy 604 VDFLVQNSGMLDKOTGKSTALHYCCLTDNAECUKLLRGKASIEIANESGETPLDIANKR 663
Db 432 VDFLVQNSGMLDKOTGKSTALHYCCLTDNAECUKLLRGKASIEIANESGETPLDIANKR 491
Qy 664 LKHEHCELLTQALSGRFSNHSVHVEYEWRLHEDLDESDDD 704
Db 492 LKHEHCELLTQALSGRFSNHSVHVEYEWRLHEDLDESDDD 532

RESULT 3
Q66JN2
ID Q66JN2 PRELIMINARY; PRT; 809 AA.
AC Q66JN2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Gene model 592.
GN Name=Gm592;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CS7BL/6; TISSUE=Brain;
RX  PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heideh F.,
RA  Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA  Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA  Jones S.J., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length human
RT  and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CS7BL/6; TISSUE=Brain;
RX  Director MGC Project;
RL  Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
CC  -1- SIMILARITY: Contains 1 SH3 domain.
DR  EMBL; BC080847; AAR80847.1; -.
DR  InterPro; IPR002110; ANK.
DR  InterPro; IPR000108; Neu_cyt_fact_2.
DR  InterPro; IPR011036; PH_related.
DR  InterPro; IPR001452; SH3.
DR  InterPro; IPR011511; SH3_2.
DR  Pfam; PF00023; Ank; 2.
DR  Pfam; PF00018; SH3_1; 1.
DR  Pfam; PF07653; SH3_2; 1.
DR  PRINTS; PR01415; ANKYRIN.
DR  PRODOM; PD000066; SH3; 1.
DR  SMART; SM00248; ANK; 3.
DR  SMART; SM00105; ArfGap; 1.
DR  SMART; SM00233; PH; 1.
DR  SMART; SM00326; SH3; 1.
DR  PRINTS; PR00499; P67PHOX.
DR  PRODOM; PD000066; SH3; 1.
DR  SMART; SM00248; ANK; 3.
DR  SMART; SM00105; ArfGap; 1.
DR  PROSITE; PS50088; ANK_REPEAT; 1.
DR  PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR  PROSITE; PS50003; PH_DOMAIN; 1.
DR  PROSITE; PS50002; SH3; 1.
DR  ANK repeat; SH3 domain.
KW  ANK repeat; SH3 domain.
SQ  SEQUENCE 809 AA; 90021 MW; B77505164F392CCA CRC64;

Query Match 16.0%; Score 161; DB 2; Length 809;
Best Local Similarity 100.0%; Pred. No. 4.1e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 603
Db 398 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 457

Qy 604 VDFLVQNSGNLDKQTGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 663
Db 458 VDFLVQNSGNLDKQTGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 517

Qy 664 LKHEHCELLTQALSGRNSHVHVEYEWRLHEDLDESDDD 704
Db 518 LKHEHCELLTQALSGRNSHVHVEYEWRLHEDLDESDDD 558

RESULT 4
Q6A074 PRELIMINARY; PRT; 970 AA.
AC Q6A074;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE MKTAA0400 protein (Fragment).
GN Mus musculus (Mouse).
OS Eukaryota; Euteleostomi;
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Okazaki N., Kikuno R.P., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-Fractionated Libraries.";
RL DNA Res. 11:205-218 (2004).
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AK172944; BA032222.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGap.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3_2.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGap; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00105; ArfGap; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ArfGap; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
DR ANK repeat; SH3 domain.
KW ANK repeat; SH3 domain.
SQ SEQUENCE 970 AA; 107768 MW; 11BAA16234DFEBC1 CRC64;

Query Match 16.0%; Score 161; DB 2; Length 970;
Best Local Similarity 100.0%; Pred. No. 4.8e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 603
Db 559 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTSLHI 618

Qy 604 VDFLVQNSGNLDKQTGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 663
Db 619 VDFLVQNSGNLDKQTGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 678

Qy 664 LKHEHCELLTQALSGRNSHVHVEYEWRLHEDLDESDDD 704
Db 679 LKHEHCELLTQALSGRNSHVHVEYEWRLHEDLDESDDD 719

RESULT 5
Q8N282 PRELIMINARY; PRT; 208 AA.
AC Q8N282;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

```

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ33802.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Tongue;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musashino K., Yuki H., Oshima A., Sasaki N., Rotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Moniyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsuura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45 (2004).  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
DR EMBL; AK091121; BAC03588.1; -.  
DR HSSP; Q9NZM3; 1UFF.  
DR InterPro; IPR000108; Neu\_cyt\_fact\_2.  
DR Pfam; PF00018; SH3\_1; 1.  
DR PRINTS; PR00499; P67PHOX.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW SH3 domain.  
SQ SEQUENCE 208 AA; 21869 MW; 8ABA541403C9986F CRC64;  
  
Query Match 10.9%; Score 110; DB 2; Length 208;  
Best Local Similarity 100.0%; Pred. No. 2.2e-103;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 897 DKSTPLTKNGQPRGPDLSATEALGPLSNAMVLPAPMPKRSQATKLPKRKVALYNCV 956  
DQ |||||||  
Db 99 DKSTPLTKNGQPRGPDLSATEALGPLSNAMVLPAPMPKRSQATKLPKRKVALYNCV 158  
DQ |||||||  
QY 957 ADNPDELTFSEGDVLIIVDGEEDQEWIIGHIDGPRKGAFVSVFVHFIAD 1006  
DQ |||||||  
Db 159 ADNPDELTFSEGDVLIIVDGEEDQEWIIGHIDGPRKGAFVSVFVHFIAD 208  
DQ |||||||  
  
RESULT 6  
Q6NRKO PRELIMINARY; PRT; 1029 AA.  
ID Q6NRKO  
AC Q6NRKO  
DC 05-JUL-2004 (TReMBLrel. 27, Created)

DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DE MG83760 protein.  
GN Name=MG83760;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosa S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Small U., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative";  
RL Dev. Dyn. 225:384-391 (2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocytes;  
RA Klein S., Strausberg R.;  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
DR EMBL; BC070750; AAH70750.1; -.  
DR HSSP; P19878; 1K4U.  
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR001164; ArfGAP.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR011036; PH related.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00023; Ank; 2.  
DR Pfam; PF01412; ArfGAP; 1.  
DR Pfam; PF00169; PH; 1.  
DR PRINTS; PR01415; ANKYRIN.  
DR PRINTS; PR00405; REVINTRACTNG.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00105; ArfGAP; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 2.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50115; ARFGAP; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW ANK repeat; SH3 domain.

```

SQ SEQUENCE 1029 AA; 114129 MW; BC5FCS7D0044A8C5 CRC64;
Query Match 4.6%; Score 46; DB 2; Length 1029;
Best Local Similarity 100.0%; Pred. No. 3.6e-37;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 194 FOLQCEYLLKVKNEIKIKGVLDLQNLKIKYFHAQCNFFQDGLKAVE 239
Db 194 FOLQCEYLLKVKNEIKIKGVLDLQNLKIKYFHAQCNFFQDGLKAVE 239

RESULT 7
ID Q8K115 PRELIMINARY; PRT; 108 AA.
AC Q8K115;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 9030624G23 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC028949; AAH28949.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; K2AB.
DR Pfam; PF01352; K2AB; 1.
DR SMART; SM00349; K2AB; 1.
DR PROSITE; PS00805; K2AB; 1.
SQ SEQUENCE 108 AA; 12545 MW; B455B44B3B24C2C CRC64;

Query Match 3.4%; Score 34; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 9.4e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAOCR 34
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAOCR 34

RESULT 8
DDFL_BOVIN

```



DR InterPro; IPR001849; PH.  
 DR InterPro; IPR011036; PH related.  
 DR InterPro; IPR011452; SH3.  
 DR Pfam; PF00023; Ank; 2.  
 DR Pfam; PF01412; ArfGap; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00499; P67PHOX.  
 DR PRINTS; PR00405; REINTRACTNG.  
 DR ProDom; PD00066; SH3; 1.  
 DR SMART; SM00248; Ank; 2.  
 DR SMART; SM00105; ArfGap; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50297; ANK REP REGION; 1.  
 DR PROSITE; PS50088; ANK REPEAT; 2.  
 DR PROSITE; PS50115; ARFGAP; 1.  
 DR PROSITE; PS50003; PH DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW ANK repeat; Direct protein sequencing; GTPase activation; Membrane;  
 FT Repeat; SH3 domain; Zinc-finger.  
 FT DOMAIN 327 419 PH.  
 FT DOMAIN 442 565 Arf-CAP.  
 FT ZN FING 457 480 C4-type.  
 FT REPEAT 603 635 ANK 1.  
 FT REPEAT 639 668 ANK 2.  
 FT DOMAIN 786 1059 Pro-rich.  
 FT DOMAIN 1067 1129 SH3.  
 FT CONFLICT 65 65 D -> T (in Ref. 2; AA sequence).  
 SQ SEQUENCE 1129 AA; 125381 MW; C1576CEACQACDAB CRC64;  
 Query Match 3.0%; Score 30; DB 1; Length 1129;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-21;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 127 LDSLLKGLKGVKGLKPKFPKAWKDYETK 156  
 |||||  
 Db 147 LDSLLKGLKGVKGLKPKFPKAWKDYETK 176  
 |||||  
 RESULT 9  
 DDF1\_MOUSE STANDARD; PRT; 1147 AA.  
 AC Q9QWY8; O08612; Q80TC8; Q80UV6; Q99LV8; Q922B6;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-  
 DE activating protein (PIP2-dependent ARF1 GAP) (ADP-ribosylation factor-  
 DE directed GTPase-activating protein 1) (ARF GTPase-activating protein  
 DE 1) (Development and differentiation-enhancing factor 1)  
 DE (Differentiation-enhancing factor 1) (DEF-1).  
 GN Name=Def1; Synonyms=Asap1, Kiaa1249, Shag1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND  
 RP MUTAGENESIS OF ARG-811; PRO-910 AND PRO-913.  
 RC TISSUE=Brain, and Embryo;  
 RX MEDLINE=99038209; PubMed=9819391;  
 RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,  
 RA Randazzo P.A.;  
 RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that  
 RT associates with and is phosphorylated by Src.";  
 RL Mol. Cell. Biol. 18:7038-7051(1998).  
 RN [2]  
 RP SEQUENCE OF 57-1147 FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=22579291; PubMed=13693553;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
 RA Nakajima D., Nagase T., Ohara O., Koga H.;  
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
 RT I1. The complete nucleotide sequences of 400 mouse KIAA-homologous  
 RT cDNAs identified by screening of terminal sequences of cDNA clones  
 RT randomly sampled from size-fractionated libraries.";  
 RL DNA Res. 10:35-48(2003).  
 RN [3]  
 RP SEQUENCE OF 654-1147 FROM N.A. (ISOFORM 1).  
 RX MEDLINE=97271433; PubMed=9126384; DOI=10.1006/abio.1997.2040;  
 RA Yamabhai M., Kay B.K.;  
 RT "Examining the specificity of Src homology 3 domain -- ligand  
 RT interactions with alkaline phosphatase fusion proteins.";  
 RL Anal. Biochem. 247:143-151(1997).  
 RN [4]  
 RP SEQUENCE OF 20-1147 FROM N.A. (ISOFORM 4), AND SEQUENCE OF 722-1147  
 RP FROM N.A. (ISOFORM 1).  
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
 RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Smerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=99147067; PubMed=10022919;  
 RA King P.J., Hu E., Harris D.F., Sarraf P., Spiegelman B.M.,  
 RA Roberts T.M.;  
 RT "DEF-1, a novel src SH3 binding protein that promotes adipogenesis in  
 RT fibroblastic cell lines.";  
 RL Mol. Cell. Biol. 19:2330-2337(1999).  
 CC -1- FUNCTION: May function as a signal transduction protein involved  
 CC in the differentiation of fibroblasts into adipocytes and possibly  
 CC other cell types (by similarity). Possesses phosphatidylinositol  
 CC 4,5-bisphosphate-dependent GTPase-activating protein activity for  
 CC ARF1 (ADP ribosylation factor 1) and ARF5 and a lesser activity  
 CC towards ARF6. May coordinate membrane trafficking with cell growth  
 CC or actin cytoskeleton remodeling by binding to both SRC and PIP2.  
 CC -1- ENZYME REGULATION: Activity stimulated by phosphatidylinositol  
 CC 4,5-bisphosphate (PIP2).  
 CC -1- SUBUNIT: Homodimer. Interacts with SRC and CRK.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-  
 CC associated; partially.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=4;  
 CC Name=1; Synonyms=SHAG1a, ASAP1a;  
 CC IsoId=Q9QWY8-1; Sequence=Display;  
 CC Name=2; Synonyms=SHAG1b, ASAP1b;  
 CC IsoId=Q9QWY8-2; Sequence=VSP\_008368;  
 CC Name=3;  
 CC IsoId=Q9QWY8-3; Sequence=VSP\_008366;  
 CC Name=4;  
 CC IsoId=Q9QWY8-4; Sequence=VSP\_008367;  
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined but a most  
 CC abundant expression was found in the testis, brain, lung and  
 CC spleen. A heightened expression was seen in the adipose tissue  
 CC from obese (ob) and diabetic (db) animals.  
 CC -1- DOMAIN: The PH domain most probably contributes to the  
 CC phosphoinositide-dependent regulation of ADP ribosylation factors.

CC -1- PTM: Phosphorylated on tyrosine residues by SRC.  
 CC -1- SIMILARITY: Contains 2 ANK repeats.  
 CC -1- SIMILARITY: Contains 1 Arf-GAP domain.  
 CC -1- SIMILARITY: Contains 1 C4-type zinc finger.  
 CC -1- SIMILARITY: Contains 1 PH domain.  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF075461; AAC98349.1; -;  
 CC EMBL; AF075462; AAC98350.1; -;  
 CC EMBL; AK122477; BAC65759.1; -;  
 CC EMBL; BC002201; AAH02201.1; ALT INIT.  
 CC EMBL; BC048818; AAH48818.1; ALT\_INIT.  
 CC EMBL; U92478; AAB82338.1; -;  
 CC PIR; T42627; T42627.  
 CC HSSP; P15891; IJ08.  
 CC MGD; MGI:1342335; Ddef1.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR001164; hRIP-like.  
 CC InterPro; IPR00108; New\_Cyt\_fact\_2.  
 CC InterPro; IPR001849; PH.  
 CC InterPro; IPR011036; PH related.  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00023; Ank; 3.  
 CC Pfam; PF01412; ArfGap; 1.  
 CC Pfam; PF00169; PH; 1.  
 CC Pfam; PF00018; SH3; 1.  
 CC PRINTS; PR00499; P67PHOX.  
 CC PRINTS; PR00405; REVINTRACTING.  
 CC ProDom; PD000066; SH3; 1.  
 CC SMART; SM00248; ANK; 2.  
 CC SMART; SM00105; ArfGap; 1.  
 CC SMART; SM00233; PH; 1.  
 CC SMART; SM00326; SH3; 1.  
 CC PROSITE; PS0297; ANK REP REGION; 1.  
 CC PROSITE; PS00088; ANK\_REPEAT; 2.  
 CC PROSITE; PS00115; ARFGAP; 1.  
 CC PROSITE; PS00003; PH DOMAIN; 1.  
 CC PROSITE; PS00002; SH3; 1.  
 CC Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;  
 CC SH3 domain; zinc-finger.  
 CC FT DOMAIN 339 431 PH.  
 CC FT DOMAIN 454 577 Arf-CAP.  
 CC FT ZN\_FING 469 492 C4-type.  
 CC FT REPEAT 615 647 ANK 1.  
 CC FT REPEAT 651 680 ANK 2.  
 CC FT DOMAIN 798 1011 Pro-rich.  
 CC FT DOMAIN 1085 1147 SH3.  
 CC FT VARSPPLIC 304 315 Missing (in isoform 3).  
 CC FT VARSPPLIC 304 318 Missing (in isoform 4).  
 CC FT VARSPPLIC 304 318 Missing (in isoform 4).  
 CC FT VARSPPLIC 816 872 Missing (in isoform 2).  
 CC FT VARSPPLIC 816 872 Missing (in isoform 2).  
 CC FT MUTAGEN 811 811 R->A: Significant reduction in binding to  
 CC SRC and CRK and loss of phosphorylation.  
 CC FT SRC and CRK and loss of phosphorylation.  
 CC FT Loss of binding and phosphorylation; when  
 CC associated with A-910 and A-913.  
 CC FT P->A: Significant reduction in binding to  
 CC SRC and CRK and decrease in  
 CC phosphorylation; when associated with A-  
 CC 913. Loss of binding and phosphorylation;  
 CC when associated with A-811 and A-913.  
 CC FT P->A: Significant reduction in binding to  
 CC SRC and CRK and decrease in  
 CC phosphorylation; when associated with A-  
 CC 913.  
 CC FT MUTAGEN 913 913

FT FT 910. Loss of binding and phosphorylation;  
 FT when associated with A-811 and A-910.  
 FT T -> S (in Ref. 4).  
 FT S -> L (in Ref. 3 and 4).  
 FT R -> I (in Ref. 4).  
 SQ SEQUENCE 1147 AA; 127395 MW; 1A08321C491B4609 CRC64;  
 Query Match 3.08; Score 30; DB 1; Length 1147;  
 Best Local Similarity 100.0%; Pred. NO. 9.8e-21;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 127 LDSLLKGLDVGKVGDLKKPKDKAWKDYETK 156  
 |||||  
 Db 147 LDSLLKGLDVGKVGDLKKPKDKAWKDYETK 176  
 |||||  
 RESULT 10  
 Q8BYB8 PRELIMINARY; PRT; 111 AA.  
 ID Q8BYB8  
 AC Q8BYB8  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched  
 DE library, clone:A630001111 product:K1AA0400 PROTEIN homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RT Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20499174; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RA "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";

```

RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayaishizaki Y.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041326; BAC30907.1; -
SQ SEQUENCE 111 AA; 11591 MW; 551D3D2E51983077 CRC64;

Query Match 2.8%; Score 28; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.4e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPDQISVSSEFVAETHEDYKAPTASFTT 28
Db 1 MPDQISVSSEFVAETHEDYKAPTASFTT 28

RESULT 11
ID Q8K0U0 PRELIMINARY; PRT; 98 AA.
AC Q8K0U0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN CDNA 903062423 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=2328257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villaion D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC030401; AAH30401.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

```

```

DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS08005; KRAB; 1.
SQ SEQUENCE 98 AA; 10971 MW; 3BF7702A67E3535A CRC64;

Query Match 2.3%; Score 23; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.6e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 APTASSFTTRTAQCRNTVAALIEE 42
Db 20 APTASSFTTRTAQCRNTVAALIEE 42

RESULT 12
ID DDF1 HUMAN STANDARD; PRT; 956 AA.
AC Q9ULH1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-
DE activating protein (PIP2-dependent ARF1 GAP) (ADP-ribosylation factor-
DE directed GTPase-activating protein 1) (ARF GTPase-activating protein
DE 1) (Development and differentiation-enhancing factor 1) (Fragment).
DE Name=DDF1; Synonyms=ASAP1, KIAA1249;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Amid C., Oanger A.,
RA Fobos G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 11-959 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20035619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345 (1999).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=99038209; PubMed=9819391;
RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
RA Randazzo P.A.;
RT ASAP1, a phospholipid-dependent arf GTPase-activating protein that
RT associates with and is phosphorylated by Src."
RL Mol. Cell. Biol. 18:7038-7051 (1998).
CC -1- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent
CC GTPase-activating protein activity for ARF1 (ADP ribosylation
CC factor 1) and ARF5 and a lesser activity towards ARP6. May
CC coordinate membrane trafficking with cell growth or actin
CC cytoskeleton remodeling by binding to both SRC and PIP2. May
CC function as a signal transduction protein involved in the
CC differentiation of fibroblasts into adipocytes and possibly other
CC cell types (By similarity).
CC -1- ENZYME REGULATION: Activity stimulated by phosphatidylinositol
CC 4,5-bisphosphate (PIP2) (By similarity).
CC -1- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-
CC associated; partially (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2;
CC IsoId=Q9ULH1-1; Sequence=Displayed;
CC Name=1;

```

```

CC CC      IsoId=Q9ULH1-2; Sequence=VSP_008365;
CC CC      -!- DOMAIN: The PH domain most probably contributes to the
CC CC      phosphoinositide-dependent regulation of ADP ribosylation factors
CC CC      (by similarity).
CC CC      -!- PTM: Phosphorylated on tyrosine residues by SRC (by similarity).
CC CC      -!- SIMILARITY: Contains 2 ANK repeats.
CC CC      -!- SIMILARITY: Contains 1 Arf-GAP domain.
CC CC      -!- SIMILARITY: Contains 1 C4-type zinc finger.
CC CC      -!- SIMILARITY: Contains 1 PH domain.
CC CC      -!- SIMILARITY: Contains 1 SH3 domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; BX537768; CAD97831.1; -;
DR DR      EMBL; AB033075; BAA86563.1; -;
DR DR      InterAct; Q9ULH1; -;
DR DR      Genew; HGNC:2720; DDEF1.
DR DR      MIN; 605953; -;
DR DR      InterPro; IPR002110; ANK.
DR DR      InterPro; IPR001164; hrip like.
DR DR      InterPro; IPR000108; Neu_Cyt_fact_2.
DR DR      InterPro; IPR001849; PH.
DR DR      InterPro; IPR011036; PH_related.
DR DR      InterPro; IPR001452; SH3.
DR DR      Pfam; PF00023; ANK; 2.
DR DR      Pfam; PF01412; ArfGAP; 1.
DR DR      Pfam; PF00169; PH; 1.
DR DR      Pfam; PF00018; SH3; 1.
DR DR      PRINTS; PR00499; P67PHOX.
DR DR      PRINTS; PR00405; REVINTRACTNG.
DR DR      PRINTS; PR00452; SH3DOMAIN.
DR DR      ProDom; PD000066; SH3; 1.
DR DR      SMART; SM00248; ANK; 2.
DR DR      SMART; SM00105; ArfGAP; 1.
DR DR      SMART; SM00233; PH; 1.
DR DR      SMART; SM00326; SH3; 1.
DR DR      PROSITE; PS50297; ANK REP REGION; 1.
DR DR      PROSITE; PS50088; ANK REPEAT; 1.
DR DR      PROSITE; PS50115; ARFGAP; 1.
DR DR      PROSITE; PS50003; PH_DOMAIN; 1.
DR DR      PROSITE; PS50002; SH3; 1.
DR DR      KW      Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;
DR DR      SH3 domain; Zinc-finger.
FT FT      NON TER      1      243      PH
FT FT      DOMAIN      151      243      Arf-GAP.
FT FT      ZN_FING      281      304      C4-type.
FT FT      REPEAT      427      459      ANK 1.
FT FT      REPEAT      463      492      ANK 2.
FT FT      DOMAIN      610      820      Pro-rich.
FT FT      DOMAIN      894      956      SH3.
FT FT      VARSP LIC      130      130      E -> ESRR (in isoform 1).
FT FT      /FTID=VSP_008365.
SQ      SEQUENCE      956 AA; 105967 MW; 6BCBB7034EC5BBE0 CRC64;
Query Match      2.2%; Score 22; DB 1; Length 956;
Best Local Similarity      100.0%; Pred. No. 1.3e-12;
Matches      22; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
Qy      444      PTLSTNLGILTCICSGIHRE      465
Db      289      PTLSTNLGILTCICSGIHRE      310
RESULT 13
Q8C1G6
ID      Q8C1G6      PRELIMINARY;      PRT;      123 AA.

```

```

AC      Q8C1G6;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
DE      enriched library, clone:6030426L16 product:similar to KRUPPEL-RELATED
DE      ZINC FINGER PROTEIN F80-L.
GN      Name=6030426L16Rik;
OS      Mus musculus (Mouse);
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RX      MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RT      Carninci P., Hayashizaki Y.;
RT      "High-efficiency full-length cDNA cloning.";
RL      Meth. Enzymol. 303:19-44(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RX      MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RT      RIKEN FANTOM Consortium;
RT      "Functional annotation of a full-length mouse cDNA collection.";
RL      Nature 409:685-690(2001).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RT      The FANTOM Consortium;
RA      "Analysis of the mouse transcriptome based on functional annotation of
RT      60,770 full-length cDNAs.";
RL      Nature 420:563-573(2002).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RX      MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA      Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA      Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT      "Normalization and subtraction of cap-trapper-selected cDNAs to
RT      prepare full-length cDNA libraries for rapid discovery of new genes.";
RL      Genome Res. 10:1617-1630(2000).
RN      [5]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RX      MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA      Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA      Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA      Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA      Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishwagi K.,
RA      Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA      Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA      Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT      "RIKEN integrated sequence analysis (RISA) system-384-format
RT      sequencing pipeline with 384 multicapillary sequencer.";
RL      Genome Res. 10:1757-1771(2000).
RN      [6]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CS7BL/6J; TISSUE=Testis;
RA      Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA      Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA      Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA      Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA      Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA      Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA      Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA      Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA      Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA      Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA      Muramatsu M., Hayashizaki Y.;
RA      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL      EMBL; AK020054; BAC25612.1; -.

```

```

DR MGD; MGI:1924364; 6030426L16Rik.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB; 1
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS08085; KRAB; 1.
SQ SEQUENCE 123 AA; 14105 MW; 445173CF14A1A3AE CRC64;

Query Match 1.9%; Score 19; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 EDYKAPTASSFTTTRTAQCR 34
Db 16 EDYKAPTASSFTTTRTAQCR 34

RESULT 14
ID Q8RL15 PRELIMINARY; PRT; 121 AA.
AC Q8RL15;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN CDNA 9030624G23 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024416; AAH24416.1;
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS08085; KRAB; 1.
SQ SEQUENCE 121 AA; 13904 MW; EC20C7A61B0634D1 CRC64;

Query Match 1.6%; Score 16; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MPDQISVSEFVAETHE 16
Db 1 MPDQISVSEFVAETHE 16

RESULT 15
ID Q9D308 PRELIMINARY; PRT; 155 AA.
AC Q9D308;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult male colon cDNA, RIKEN full-length enriched
DE library, clone:9030624G23 product:hypothetical protein, full insert
DE sequence.
GN Name=9030624G23Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RA Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.153600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplex capillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,

```

RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.,  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK018572; BAB31284.1; -  
DR MGD; MGI:1914058; 9030624G23Rik.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003676; P:nucleic acid binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001909; KRAB.  
DR Pfam; PF01352; KRAB; 1  
DR SMART; SM00349; KRAB; 1.  
DR PROSITE; PS50805; KRAB; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 155 AA; 17560 MW; D63034719828920E CRC64;

Query Match 1.5%; Score 15; DB 2; Length 155;  
Best Local Similarity 100.0%; Pred.No. 3.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 APTASSFTTTRTAQCR 34  
|||||  
Db 20 APTASSFTTTRTAQCR 34

Search completed: August 5, 2005, 04:05:23  
Job time : 121 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 02:52:18 ; Search time 31 Seconds  
(without alignments)  
3122.389 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 1006  
Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAPVSVVFIAD 1006

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	703	69.9	1006	2 T00050	hypothetical prote
2	30	3.0	1147	2 T42627	ADP-ribosylation f
3	12	1.2	1243	2 J55615	membrane-associate
4	10	1.0	138	2 G55096	hypothetical prote
5	9	0.9	903	2 T26743	hypothetical prote
6	9	0.9	923	2 T38398	hypothetical GTPas
7	9	0.9	951	2 T26738	hypothetical prote
8	9	0.9	975	2 T26737	hypothetical prote
9	9	0.9	1087	2 T49496	hypothetical prote
10	8	0.8	153	2 H75003	hypothetical prote
11	8	0.8	155	2 D82755	conserved hypothe
12	8	0.8	185	1 S60942	hypothetical prote
13	8	0.8	220	2 T26434	hypothetical prote
14	8	0.8	236	2 H87976	probable RING zinc
15	8	0.8	237	2 G84678	hypothetical prote
16	8	0.8	285	2 A75074	hypothetical prote
17	8	0.8	312	1 JQ1559	polyhedron envelop
18	8	0.8	313	2 T30486	polyhedral calyx p
19	8	0.8	320	2 G96714	hypothetical prote
20	8	0.8	359	2 S14283	transcription fact
21	8	0.8	398	2 T41600	probable pre-mRNA
22	8	0.8	404	2 C69170	UDP-N-acetylmutam
23	8	0.8	462	2 E70955	hypothetical prote
24	8	0.8	482	2 E97748	virB10 protein (im
25	8	0.8	484	2 T07675	cyclin a2-type, mi
26	8	0.8	539	2 D71260	hypothetical prote
27	8	0.8	596	2 T03908	hypothetical prote
28	8	0.8	627	2 T04562	hypothetical prote
29	8	0.8	678	2 T40362	conserved hypothe

30	8	0.8	822	2 T01622	probable salt-indu
31	8	0.8	847	2 A82296	hypothetical prote
32	8	0.8	1292	2 D84727	probable RAD50 DNA
33	8	0.8	1948	2 B69511	N conserved hypoth
34	8	0.8	2100	2 T03223	probable polyketid
35	8	0.8	3712	2 S18253	laminin alpha-1 ch
36	8	0.8	4613	2 T17409	polyketide synthas
37	7	0.7	27	2 I57715	probasin - rat (fr
38	7	0.7	56	2 A98162	hypothetical prote
39	7	0.7	103	2 I37454	gene HE2 protein -
40	7	0.7	106	2 T00722	hypothetical prote
41	7	0.7	111	2 A53221	acidic ribosomal p
42	7	0.7	119	2 C40513	hypothetical prote
43	7	0.7	123	1 C44212	structural protein
44	7	0.7	123	2 G81127	hypothetical prote
45	7	0.7	127	2 A83861	aspartate 1-decarb

ALIGNMENTS

RESULT 1

T00050  
hypothetical protein KIAA0400 - human  
C:Species: Homo sapiens (man)  
C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00050  
R:Ichikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka, A.,  
submitted to the EMBL Data Library, October 1997  
A:Description: Prediction of the coding sequences of unidentified human genes. VIII. The  
A:Reference number: Z14080  
A:Accession: T00050  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1006 <ISH>  
A:Cross-references: EMBL:AB007860; NID:d1175316; PIDN:BA03696.1; PID:d1024577  
A:Experimental source: brain, clone HG1091  
C:Genetics:  
A:Note: KIAA0400

Query Match		69.9%;	Score 703;	DB 2;	Length 1006;
Best Local Similarity		99.7%;	Pred. No. 0;		
Matches 1003;		Conservative	0;	Mismatches	3;
				Indels	0;
				Gaps	0;
Qy	1	MPDQISVSEFVAETHDYKAPTASSFTRTAQCRNTVAAIEEALDVRMVLVYMKKSKVKA	60		
Db	1	MPDQISVSEFVAETHDYKAPTASSFTRTAQCRNTVAAIEEALDVRMVLVYMKKSKVKA	60		
Qy	61	INSSGLAHVNEEQYTALEKPGGNCVCRDDPDGLGSAPLKFSVFTKELTALFKNLIQNN	120		
Db	61	INSSGLAHVNEEQYTALEKPGGNCVCRDDPDGLGSAPLKFSVFTKELTALFKNLIQNN	120		
Qy	121	NIISFPDLSLLKGDLKGVKGLKPPDKAWDYTKTKIEKEKEHAHKLHGMRTETLSG	180		
Db	121	NIISFPDLSLLKGDLKGVKGLKPPDKAWDYTKTKIEKEKEHAHKLHGMRTETLSG	180		
Qy	181	ABIAEEMEKEKRRFQLOWCEYLLKKNVETIKKGVDDLQNLIKYFHAQCNFFQDGLKAVES	240		
Db	181	ABIAEEMEKEKRRFQLOWCEYLLKKNVETIKKGVDDLQNLIKYFHAQCNFFQDGLKAVES	240		
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDILKSALQVEQKESQIRQSTAYSLSHQPOG	300		
Db	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDILKSALQVEQKESQIRQSTAYSLSHQPOG	300		
Qy	301	NKEHGTNRNGSLYKKSQIRKQWQKRCVQNGFLTISHGTANRPPAKNLNLTTCQVKTNP	360		
Db	301	NKEHGTNRNGSLYKKSQIRKQWQKRCVQNGFLTISHGTANRPPAKNLNLTTCQVKTNP	360		
Qy	361	EEKKCFDLISHDRYTHFQAQDEQECQIWMVSLQNSKEEALNNAFGKDDNTGNNIVQSLT	420		
Db	361	EEKKCFDLISHDRYTHFQAQDEQECQIWMVSLQNSKEEALNNAFGKDDNTGNNIVQSLT	420		
Qy	421	KEIISVQRMGTGNDVCCDGPDPDTWLSTNLIGILTICBSCGIHRELGVHYSPMQSLTLDV	480		

```
|||||
421 KEIISVQRMGTNDVCCDGPDPWLTSTNLGILTCIEGSHRELGVHYSRMQSLTLDV 480
Qy LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKQDYITAKYTIERYARKKH 540
Db LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKQDYITAKYTIERYARKKH 540
Qy ADNAKHLSCAEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVSVORTS 600
Db ADNAKHLSCAEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVSVORTS 600
Qy LHIVDLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db LHIVDLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMKQPSNRRDRP 720
Db AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMKQPSNRRDRP 720
Qy ISFYOLGSOIQSNAVSLARDAANLAKKORAFWPSILQNETYGNLLSGSPPPQAAPS 780
Db ISFYOLGSOIQSNAVSLARDAANLAKKORAFWPSILQNETYGNLLSGSPPPQAAPS 780
Qy TTSAPPLPPRVNVKQVTASSANTLWKTNSVSDGGRQSSSDPPAVHPPLPLRLVSTN 840
Db TTSAPPLPPRVNVKQVTASSANTLWKTNSVSDGGRQSSSDPPAVHPPLPLRLVSTN 840
Qy PLTTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPPOPPSRLPQKXPAGTKST 900
Db PLTTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPPOPPSRLPQKXPAGTKST 900
Qy PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Db PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Qy DELTFSEGDVIIVDGEEDQEWIGHIDGDPGRKGAFFVSVFHFAD 1006
Db DELTFSEGDVIIVDGEEDQEWIGHIDGDPGRKGAFFVSVFHFAD 1006

RESULT 2
T42627
ADP-ribosylation factor-directed GTPase activating protein, isoform a - mouse
N:Alternate names: ASAP1a protein
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42627
R:Brown, M.T.; Andrade, J.; Radhakrishna, H.; Donaldson, J.G.; Cooper, J.A.; Randazzo, F.
Mol. Cell. Biol. 18, 7038-7051, 1998
A:Title: ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates w
A:Reference number: 222178; MUID:99038209; PMID:9819391
A:Accession: T42627
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1147 <BRO>
A:Cross-references: UNIPROT:Q9QWY8; EMBL:AF075461; NID:g4063613; PID:g4063614; PIDN:AAC9
C:Genetics:
A:Gene: Shag1
C:Function:
A:Description: interacts directly with ADP-ribosylation factors (Arf1, Arf5 and to a les
A:Note: supposed to coordinate membrane remodeling events
C:Keywords: membrane trafficking

Query Match 3.0%; Score 30; DB 2; Length 1147;
Best Local Similarity 100.0%; Pred. No. 1.6e-21;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy LDSLLKGDLLKGVKGDLLKXPFDPKAWKDYEYTK 156
Db LDSLLKGDLLKGVKGDLLKXPFDPKAWKDYEYTK 176

RESULT 3
```

```
JC5615
membrane-associated phosphatidyl inositol transfer protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 09-Jul-2004
C:Accession: JC5615
R:Aikawa, Y.; Hara, H.; Watanabe, T.
Biochem. Biophys. Res. Commun. 236, 559-564, 1997
A:Title: Molecular cloning and characterization of mammalian homologues of the Drosophil
A:Reference number: JC5615; MUID:97396134; PMID:9245688
A:Contents: Brain
A:Accession: JC5615
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1243 <AIK>
A:Cross-references: UNIPROT:O35954; DDBJ:AF006467
C:Comment: This protein plays a role in brain development, at a particular stage by tran
C:Genetics:
A:Gene: mpt-1
A:Map position: 19

Query Match 1.2%; Score 12; DB 2; Length 1243;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 772 PPAQPAAPSTTS 783
Db 811 PPAQPAAPSTTS 822

RESULT 4
G65096
hypothetical protein ygjM - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: G65096
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65096
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-138 <BLAT>
A:Cross-references: UNIPROT:P42594; GB:AE000390; GB:U00096; NID:g2367189; PIDN:AAC76117.
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: YgJm
C:Superfamily: Escherichia coli hypothetical protein ygjM

Query Match 1.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 71 NEEQYTOALE 80
Db 26 NEEQYTOALE 35

RESULT 5
T26743
hypothetical protein Y39A1A.15c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26743
R:Wall, M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z20257
A:Accession: T26743
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-903 <WIL>
A:Cross-references: UNIPROT:Q9XX09; EMBL:AL031633; PIDN:CAA21032.1; GSPDB:GN00021; CESP:
```



A:Experimental source: clone Y39A1A

C:Genetics:

A:Gene: CESP:Y39A1A.15c

A:Map position: 3

A:Introns: 33/3; 63/1; 139/3; 185/1; 382/3; 424/2; 798/3; 852/3

Query Match 0.9%; Score 9; DB 2; Length 903;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 653 CIECSGIHR 661

#### RESULT 6

T38398 hypothetical GTPase activating protein for Arf protein - fission yeast (Schizosaccharomy

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T38398

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z21791

A:Accession: T38398

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-923 <MCL>

A:Cross-references: UNIPROT:Q10165; EMBL:Z69240; PIDN:CAA93233.1; GSPDB:GN00066; SPDB:SF

A:Experimental source: strain 972h-; cosmid 26A3

C:Genetics:

A:Gene: SPDB:SPAC26A3.10

A:Map position: 1

A:Introns: 6/1; 47/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 923;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 751 CIECSGIHR 759

#### RESULT 7

T26738

hypothetical protein Y39A1A.15b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T26738

R:Wall, M.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z20257

A:Accession: T26738

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-951 <WIL>

A:Cross-references: UNIPROT:Q9U2M2; EMBL:AL031633; PIDN:CAA21027.1; GSPDB:GN00021; CESP:

A:Experimental source: clone Y39A1A

C:Genetics:

A:Gene: CESP:Y39A1A.15b

A:Map position: 3

A:Introns: 34/3; 62/1; 81/3; 111/1; 187/3; 233/1; 430/3; 472/2; 846/3; 900/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 951;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 701 CIECSGIHR 709

#### RESULT 8

T26737

hypothetical protein Y39A1A.15a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T26737

R:Wall, M.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z20257

A:Accession: T26737

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-975 <WIL>

A:Cross-references: UNIPROT:Q9XX14; EMBL:AL031633; PIDN:CAA21026.1; GSPDB:GN00021; CESP:

A:Experimental source: clone Y39A1A

C:Genetics:

A:Gene: CESP:Y39A1A.15a

A:Map position: 3

A:Introns: 105/3; 135/1; 211/3; 257/1; 454/3; 496/2; 870/3; 924/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 975;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 725 CIECSGIHR 733

#### RESULT 9

T49496

hypothetical protein B14D6.480 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49496

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49496

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1087 <SCH>

A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.480

A:Experimental source: BAC clone B14D6; strain OR74A

C:Genetics:

A:Gene: NCSP:B14D6.480

A:Map position: 6

A:Introns: 67/3; 99/3; 270/2; 589/2; 616/2

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 1087;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 811 CIECSGIHR 819

#### RESULT 10

H75003

hypothetical protein PAB1032 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C:Accession: H75003

R:Anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: H75003

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-153 <KAW>

A:Cross-references: UNIPROT:Q9UYE2; GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB5047

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1032

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB1032

Query Match 0.8%; Score 8; DB 2; Length 153;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 ITKIEKEK 164

Db 123 ITKIEKEK 130

RESULT 11

D82755

conserved hypothetical protein XF0849 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C:Accession: D82755

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82755

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <SIM>

A:Cross-references: UNIPROT:Q9PF29; GB:AE003924; GB:AE003849; NID:g9105750; PIDN:AAF8365

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.B.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

, F.G.; Nunes, L.R.; Oliveira, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0849

Query Match

Best Local Similarity 0.8%; Score 8; DB 2; Length 155;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 HGMIRTEI 178

Db 94 HGMIRTEI 101

RESULT 12

S60942

hypothetical protein YOR215c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O5025; hypothetical protein YOR50-5

C:Species: Saccharomyces cerevisiae

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C:Accession: S60942; S67108; S71717

R:Galissou, F.; Dujon, B.

submitted to the EMBL Data Library, October 1995

A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosom

A:Reference number: S60938

A:Accession: S60942

A:Molecule type: DNA

A:Residues: 1-185 <GAL>

A:Cross-references: UNIPROT:Q12032; EMBL:X92441; NID:g1050762; PID:g1050767

R:Boyer, J.; Fairhead, C.; Gaillon, L.; Galissou, F.; Michaux, G.; Thierry, A.; Dujon, B

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67104

A:Accession: S67108

A:Molecule type: DNA

A:Residues: 1-185 <BOY>

A:Cross-references: EMBL:Z75123; NID:g1420500; PID:g1420501; GSPDB:GN00015; MIPS:YOR215c

A:Experimental source: strain S288C

R:Galissou, F.; Dujon, B.

Yeast 12, 877-885, 1996

A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV o

A:Reference number: S71713; MUID:96437977; PMID:18840505

A:Accession: S71717

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-185 <GAW>

A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63178.1; PID:g1050767

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Gene: MIPS:YOR215c

A:Cross-references: SGD:S0005741

A:Map position: 1SR

C:Superfamily: Bacillus subtilis conserved hypothetical protein yqey

Query Match

Best Local Similarity 0.8%; Score 8; DB 1; Length 185;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 SLLKGDLK 136

Db 30 SLLKGDLK 37

RESULT 13

T26434

hypothetical protein Y106G6H.14 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T26434

R:McMurray, A.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20214

A:Accession: T26434

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-220 <WIL>

A:Cross-references: UNIPROT:Q9U2Z5; EMBL:AL032631; PIDN:CAA21577.2; GSPDB:GN00019; CBSP:

A:Experimental source: clone Y106G6H

C:Genetics:

A:Gene: CESP:Y106G6H.14

A:Map position: 1

A:Introns: 32/3; 72/1; 143/3; 188/3

Query Match

Best Local Similarity 0.8%; Score 8; DB 2; Length 220;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 962 ELTFSEGD 969

Db 35 ELTFSEGD 42

RESULT 14

H87976

protein Y106G6H.14 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: H87976

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/c\_elegans/ and www.sanger.ac.uk/Projects/C\_ele

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: H87976  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-236 <STO>  
A;Cross-references: UNIPROT:Q9U2Z5; GB:chr\_I; PIDN:CAA21577.1; PID:G3880686; GSPDB:GN000  
C;Genetics:  
A;Gene: Y106G6H.14  
A;Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 236;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 ELTFSEGD 969  
|||  
Db 35 ELTFSEGD 42

RESULT 15  
G84678  
Probable RING zinc finger protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: G84678  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: G84678  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-237 <STO>  
A;Cross-references: UNIPROT:Q9SUJ7; GB:AE002093; NID:G4510422; PIDN:AAD21508.1; GSPDB:GN  
C;Genetics:  
A;Gene: At2G27940  
A;Map position: 2

Query Match 0.8%; Score 8; DB 2; Length 237;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 841 PLTPTPPP 848  
|||  
Db 33 PLTPTPPP 40

Search completed: August 5, 2005, 04:05:56  
Job time : 34 secs

***THIS PAGE BLANK (USPTO)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 04:05:29 ; Search time 104 Seconds  
(without alignments)  
3776.348 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 1006  
Sequence: 1 MPDQISVSEFAETHEDEYKA.....DGDGRKGAPVSVFVFIAD 1006

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 1752860 seqs, 390397842 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1752860

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Application: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	10.9	208	US-10-104-047-2284	Sequence 2284, App
2	110	10.9	350	US-09-764-875-635	Sequence 635, App
3	79	7.9	349	US-09-764-875-956	Sequence 956, App
4	30	3.0	136	US-10-276-774-1754	Sequence 1754, App
5	30	3.0	358	US-10-408-765A-2815	Sequence 2815, App
6	30	3.0	1132	US-10-210-281-74	Sequence 74, App
7	16	1.6	156	US-10-276-774-1638	Sequence 1638, App
8	13	1.3	43	US-09-864-761-46527	Sequence 46527, A
9	10	1.0	138	US-09-912-020-300	Sequence 300, App
10	10	1.0	138	US-10-771-241-300	Sequence 300, App
11	10	1.0	903	US-10-104-047-2951	Sequence 2951, App

12	10	1.0	903	US-10-490-605-2	Sequence 2, App
13	10	1.0	1358	US-10-778-804-11	Sequence 11, App
14	10	0.9	122	US-10-080-334-240	Sequence 240, App
15	9	0.9	125	US-10-437-963-156203	Sequence 156203, App
16	9	0.9	138	US-10-176-306-44	Sequence 44, App
17	9	0.9	326	US-10-369-493-2559	Sequence 2559, App
18	9	0.9	556	US-10-424-599-22658	Sequence 22658, App
19	9	0.9	580	US-10-094-749-2012	Sequence 2012, App
20	9	0.9	641	US-10-425-115-336077	Sequence 336077, App
21	9	0.9	686	US-10-221-625-6	Sequence 6, App
22	9	0.9	726	US-10-467-434-17	Sequence 17, App
23	9	0.9	759	US-10-080-334-236	Sequence 236, App
24	9	0.9	764	US-10-080-334-80	Sequence 80, App
25	9	0.9	1778	US-10-080-334-238	Sequence 238, App
26	9	0.9	792	US-10-437-963-163744	Sequence 163744, App
27	9	0.9	894	US-10-080-334-235	Sequence 235, App
28	9	0.9	834	US-10-176-306-11	Sequence 11, App
29	9	0.9	834	US-10-080-334-237	Sequence 237, App
30	9	0.9	834	US-10-467-434-4	Sequence 4, App
31	9	0.9	836	US-10-719-993-842	Sequence 842, App
32	9	0.9	836	US-10-719-993-844	Sequence 843, App
33	9	0.9	836	US-10-719-993-844	Sequence 844, App
34	9	0.9	856	US-10-719-993-844	Sequence 845, App
35	9	0.9	870	US-10-322-281-358	Sequence 358, App
36	9	0.9	882	US-10-334-143-9	Sequence 9, App
37	8	0.8	17	US-09-880-149-61	Sequence 61, App
38	8	0.8	17	US-10-345-281-61	Sequence 61, App
39	8	0.8	17	US-10-345-281-61	Sequence 61, App
40	8	0.8	48	US-09-864-761-37629	Sequence 37629, A
41	8	0.8	59	US-10-424-599-180387	Sequence 180387, App
42	8	0.8	65	US-10-437-963-156713	Sequence 156713, App
43	8	0.8	71	US-09-349-954-21	Sequence 21, App
44	8	0.8	71	US-09-907-007-21	Sequence 21, App
45	8	0.8	71	US-10-673-708-21	Sequence 21, App

#### ALIGNMENTS

RESULT 1  
US-10-104-047-2284  
Sequence 2284, Application US/10104047  
Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20030236392A1 full length cDNA  
FILE REFERENCE: H1-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 2284  
LENGTH: 208  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-2284

Query Match 10.9%; Score 110; DB 15; Length 208;  
Best Local Similarity 100.0%; Pred. No. 2.6e-96;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 897 DKSTPLTNKQPGPVVLSATLALGPIISNMVLOPPAPMPKSOATYKTKRVRATLVNCV 956  
DB 99 DKSTPLTNKQPGPVVLSATLALGPIISNMVLOPPAPMPKSOATYKTKRVRATLVNCV 158  
QY 957 ADNPDELTFEGGVITVDEGEDEWMTGHTDGPGRGAPVSVFVFIAD 1006  
DB 159 ADNPDELTFEGGVITVDEGEDEWMTGHTDGPGRGAPVSVFVFIAD 200

RESULT 2

US-09-764-875-635  
Sequence 635, Application US/09764875  
Publication No. US20040018969A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P1202  
CURRENT APPLICATION NUMBER: US/09/764,875  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1249  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 635  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (191)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-875-635

Query Match 10.9%; Score 110; DB 11; Length 350;  
Best Local Similarity 100.0%; Pred. No. 4,2e-96;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 897 DKSTLTNKGQGRGVDSATLALGPLNANWLOPPAPMPKRSQATKLPKRVKALYNCV 956  
DB 241 DKSTLTNKGQGRGVDSATLALGPLNANWLOPPAPMPKRSQATKLPKRVKALYNCV 300  
QY 957 ADNPDELTFESGDVLIIVDGEDEQEWIIGHIDGDPKRGKAFPVSVFHFIAAD 1006  
DB 301 ADNPDELTFESGDVLIIVDGEDEQEWIIGHIDGDPKRGKAFPVSVFHFIAAD 350

RESULT 3  
US-09-764-875-956  
Sequence 956, Application US/09764875  
Publication No. US20040018969A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P1202  
CURRENT APPLICATION NUMBER: US/09/764,875  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1249  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 956  
LENGTH: 349  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (28)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (65)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (66)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (178)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (190)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (270)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-875-956

Query Match 7.9%; Score 79; DB 11; Length 349;  
Best Local Similarity 100.0%; Pred. No. 2e-66;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 928 VLOPPAPMPKRSQATKLPKRVKALYNCVADNPDELTFESGDVLIIVDGEDEQEWIIGHID 987  
DB 271 VLOPPAPMPKRSQATKLPKRVKALYNCVADNPDELTFESGDVLIIVDGEDEQEWIIGHID 330  
QY 988 GDPKRGKAFPVSVFHFIAAD 1006  
DB 331 GDPKRGKAFPVSVFHFIAAD 349

RESULT 4  
US-10-276-774-1754  
Sequence 1754, Application US/10276774  
Publication No. US20040053245A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: No. US20040053245A1 Nucleic Acids and Polypeptides  
FILE REFERENCE: 21272-030  
CURRENT APPLICATION NUMBER: US/10/276,774  
CURRENT FILING DATE: 2002-11-18  
Prior application number: 09/560,875  
Prior filing date: 2000-04-27  
Prior application number: 09/496,914  
Prior filing date: 2000-02-03  
NUMBER OF SEQ ID NOS: 2700  
SOFTWARE: Custom  
SEQ ID NO: 1754  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(136)  
OTHER INFORMATION: Xaa = any amino acid or nothing  
US-10-276-774-1754

Query Match 3.0%; Score 30; DB 15; Length 136;  
Best Local Similarity 100.0%; Pred. No. 7.4e-20;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 127 LDSLLKGDLLKGVKGDLLKPPDKAWKDYETX 156  
DB 93 LDSLLKGDLLKGVKGDLLKPPDKAWKDYETX 122

RESULT 5  
US-10-408-765A-2815  
Sequence 2815, Application US/10408765A  
Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, John D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Glenn, Gary M.  
APPLICANT: Warrick, Dale E.  
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
FILE REFERENCE: 660088.465  
CURRENT APPLICATION NUMBER: US/10/408,765A  
CURRENT FILING DATE: 2003-04-04  
NUMBER OF SEQ ID NOS: 3077  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 2815  
LENGTH: 358  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-408-765A-2815

Query Match 3.0%; Score 30; DB 16; Length 358;  
Best Local Similarity 100.0%; Pred. No. 1.8e-19;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LDSLLKGLKGVKGLKPPDKAWKDYETK 156  
Db 199 LDSLLKGLKGVKGLKPPDKAWKDYETK 228

## RESULT 6

US-10-210-281-74  
; Sequence 74, Application US/10210281  
; Publication No. US20040030096A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorman, Linda  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Zhong, Mei  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Pena, Carol E.A.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Stone, David J.  
; APPLICANT: Taupier, Raymond J., Jr.  
; APPLICANT: Casman, Stacie  
; APPLICANT: Rothenberg, Mark E.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF  
; FILE REFERENCE: 21402-4160  
; CURRENT APPLICATION NUMBER: US/10/210,281  
; CURRENT FILING DATE: 2003-02-05  
; PRIOR APPLICATION NUMBER: 60/309,501  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 60/310,291  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 60/361,775  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: 60/310,951  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/361,832  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: 60/311,292  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/311,979  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 60/312,203  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 60/313,201  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,702  
; PRIOR FILING DATE: 2001-08-20  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 191  
; SOFTWARE: CuraSeqIst version 0.1  
; SEQ ID NO 74  
; LENGTH: 1132  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-210-281-74

Query Match 3.0%; Score 30; DB 15; Length 1132;  
Best Local Similarity 100.0%; Pred. No. 4.9e-19;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LDSLLKGLKGVKGLKPPDKAWKDYETK 156

Db 147 LDSLLKGLKGVKGLKPPDKAWKDYETK 176

## RESULT 7

US-10-276-774-1638  
; Sequence 1638, Application US/10276774  
; Publication No. US20040053245A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y, Tom et al  
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 1638  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-276-774-1638

Query Match 1.6%; Score 16; DB 15; Length 156;  
Best Local Similarity 100.0%; Pred. No. 2.1e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 450 NLGILTCIECSGIHRE 465  
Db 3 NLGILTCIECSGIHRE 18

## RESULT 8

US-09-864-761-46527  
; Sequence 46527, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/235,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 46527  
 ; LENGTH: 43  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC009682.2  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
 ; OTHER INFORMATION: EST HUMAN HIT: AL044307.1, EVALUE 2.00e-08  
 ; OTHER INFORMATION: SWISSPROT HIT: Q06846, EVALUE 6.10e+00  
 ; NAME/KEY: unsure  
 ; LOCATION: 14  
 ; US-09-864-761-46527

Query Match 1.3%; Score 13; DB 9; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 0.0005; Mismatches 0; Indels 0; Gaps 0;

Qy 180 GAEIAEEMEKERR 192  
 Db 23 GAEIAEEMEKERR 35

RESULT 9  
 US-09-912-020-300  
 ; Sequence 300, Application US/09912020  
 ; Patent No. US20020045592A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Froelich, Jamie M.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
 ; FILE REFERENCE: ELITRA 001DVI  
 ; CURRENT APPLICATION NUMBER: US/09/912,020  
 ; PRIOR FILING DATE: 2001-07-23  
 ; PRIOR APPLICATION NUMBER: 09/492,709  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: 60/117,405  
 ; PRIOR FILING DATE: 1999-01-27  
 ; NUMBER OF SEQ ID NOS: 485  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 300  
 ; LENGTH: 138  
 ; TYPE: PRT  
 ; ORGANISM: E. Coli  
 ; US-09-912-020-300

Query Match 1.0%; Score 10; DB 9; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;

Qy 71 NEEQYTOALE 80  
 Db 26 NEEQYTOALE 35

RESULT 10  
 US-10-771-241-300  
 ; Sequence 300, Application US/10771241  
 ; Publication No. US2004024115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Forsyth, R. Allyn  
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
 ; FILE REFERENCE: ELITRA 001C1  
 ; CURRENT APPLICATION NUMBER: US/10/771,241  
 ; CURRENT FILING DATE: 2004-02-03  
 ; PRIOR APPLICATION NUMBER: 09/492,709  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: 60/117,405  
 ; PRIOR FILING DATE: 1999-01-27  
 ; NUMBER OF SEQ ID NOS: 485  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 300  
 ; LENGTH: 138  
 ; TYPE: PRT  
 ; ORGANISM: E. Coli  
 ; US-10-771-241-300

Query Match 1.0%; Score 10; DB 16; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;

Qy 71 NEEQYTOALE 80  
 Db 26 NEEQYTOALE 35

RESULT 11  
 US-10-104-047-2951  
 ; Sequence 2951, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
 ; FILE REFERENCE: H1-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIOR FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2951  
 ; LENGTH: 903  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-104-047-2951

Query Match 1.0%; Score 10; DB 15; Length 903;  
 Best Local Similarity 100.0%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;

Qy 359 NPEEKKCFDL 369  
 Db 356 NPEEKKCFDL 365

RESULT 12  
 US-10-490-605-2  
 ; Sequence 2, Application US/10490605  
 ; Publication No. US20050019768A1  
 ; GENERAL INFORMATION:



```
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCES, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; FILE INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATI/SEN-A0121p-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; CURRENT FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-490-605-2

Query Match 1.0%; Score 10; DB 17; Length 903;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 NPEKKKCFDL 368
Db 356 NPEKKKCFDL 365
|||||

RESULT 13
US-10-778-804-11
; Sequence 11, Application US/10778804
; Publication No. US20040224388A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Pollerman, Sarah E.
; APPLICANT: Memmott, Stephen D.
; TITLE OF INVENTION: Hyphal Growth in Fungi
; FILE REFERENCE: GC575-2
; CURRENT APPLICATION NUMBER: US/10/778,804
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: US 09/275,549
; PRIOR FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1358
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-778-804-11

Query Match 1.0%; Score 10; DB 16; Length 1358;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 839 TNPLTPTPPP 848
Db 341 TNPLTPTPPP 350
|||||

RESULT 14
US-10-080-334-240
; Sequence 240, Application US/10080334
; Publication No. US2004002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
```

```
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 240
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-240

Query Match 0.9%; Score 9; DB 15; Length 122;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 CIECSGIHR 464
Db 36 CIECSGIHR 44
|||||

RESULT 15
US-10-437-963-156203
; Sequence 156203, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 156203
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55894C.1.pep
US-10-437-963-156203
```

```
Query Match          0.9%; Score 9; DB 16; Length 125;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      876 PPPLPPQPP 884
        |||||
Db      17 PPPLPPQPP 25
```

```
Search completed: August 5, 2005, 04:16:00
Job time : 106 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: August 5, 2005, 03:45:33 ; Search time 237 Seconds  
(without alignments)  
4957.866 Million cell updates/sec

Title: US-09-914-042-1  
Perfect score: 1006  
Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAFPVSFVHFAD 1006

Scoring table: Oligo  
Gapop 60.0 , Gapext 60.0

Searched: 6959266 seqs, 1168006243 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents\_AA\_Main.\*  
1: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US10\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US11\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US12\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US13\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US14\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US15\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US16\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US17\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US18\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US19\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US20\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US21\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US22\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US23\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US24\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US25\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US26\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US27\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US28\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US29\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US30\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US31\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US32\_COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US33\_COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US34\_COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US35\_COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US36\_COMB.pep.\*  
32: /cgn2\_6/ptodata/1/paa/US37\_COMB.pep.\*  
33: /cgn2\_6/ptodata/1/paa/US38\_COMB.pep.\*  
34: /cgn2\_6/ptodata/1/paa/US39\_COMB.pep.\*  
35: /cgn2\_6/ptodata/1/paa/US40\_COMB.pep.\*  
36: /cgn2\_6/ptodata/1/paa/US41\_COMB.pep.\*  
37: /cgn2\_6/ptodata/1/paa/US42\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1006	100.0	1006	24	US-09-914-042-1	Sequence 1, Appli
2	703	69.9	1006	22	US-09-791-537-142678	Sequence 142678,
3	703	69.9	1006	24	US-09-949-002-361	Sequence 361, App
4	703	69.9	1006	24	US-09-949-003C-2272	Sequence 2272, Ap
5	693	68.9	911	24	US-09-949-002-425	Sequence 425, App
6	693	68.9	911	24	US-09-949-003C-2770	Sequence 2770, Ap
7	693	68.9	990	27	US-10-170-2058-7820	Sequence 7820, Ap
8	693	68.9	990	37	US-10-170-2058-18115	Sequence 18115, A
9	270	26.8	909	37	US-09-452-680-18114	Sequence 18114, A
10	195	19.4	826	24	US-09-914-042-2	Sequence 2, Appli
11	151	15.0	278	22	US-09-791-537-31632	Sequence 31632, A
12	110	10.9	208	27	US-10-104-047-2284	Sequence 2284, Ap
13	110	10.9	350	1	PCT-US01-01332-635	Sequence 635, App
14	110	10.9	350	22	US-09-764-875-635	Sequence 635, App
15	79	7.9	349	1	PCT-US01-01332-956	Sequence 956, App
16	79	7.9	349	22	US-09-764-875-956	Sequence 956, App
17	43	4.3	138	24	US-09-914-042-7	Sequence 7, Appli
18	39	3.9	66	37	US-09-200-109-1231	Sequence 1231, Ap
19	31	3.1	40	37	US-09-181-996-1300	Sequence 1300, Ap
20	30	3.0	115	27	US-10-170-2058-3508	Sequence 3508, Ap
21	30	3.0	136	1	PCT-US01-03800A-1754	Sequence 1754, Ap
22	30	3.0	136	28	US-10-276-774-1754	Sequence 1754, Ap
23	30	3.0	195	28	US-10-221-279-11235	Sequence 11235, A
24	30	3.0	358	30	US-10-408-765-2815	Sequence 2815, Ap
25	30	3.0	358	30	US-10-408-765A-2815	Sequence 2815, Ap
26	30	3.0	358	37	US-09-389-987-2815	Sequence 2815, Ap
27	30	3.0	358	37	US-09-412-418-2815	Sequence 2815, Ap
28	30	3.0	1072	1	PCT-US02-36151-51	Sequence 51, Appli
29	30	3.0	1090	22	US-09-791-537-127954	Sequence 127954,
30	30	3.0	1122	33	US-10-796-280-999	Sequence 999, App
31	30	3.0	1122	35	US-10-995-561-705	Sequence 705, App
32	30	3.0	1122	37	US-09-568-219-358	Sequence 358, App
33	30	3.0	1129	1	PCT-US02-36151-44	Sequence 44, Appli
34	30	3.0	1129	22	US-09-791-537-26287	Sequence 26287, A
35	30	3.0	1129	33	US-10-796-280-1000	Sequence 1000, Ap
36	30	3.0	1129	35	US-10-995-561-706	Sequence 706, App
37	30	3.0	1129	37	US-09-568-219-359	Sequence 359, App
38	30	3.0	1132	28	US-10-210-281-74	Sequence 74, Appli
39	30	3.0	1147	22	US-09-791-537-127952	Sequence 127952,
40	29	2.9	47	37	US-09-169-840-6265	Sequence 6265, Ap
41	29	2.9	54	37	US-09-170-373-3689	Sequence 3689, Ap
42	27	2.7	27	37	US-09-160-209-3848	Sequence 3848, Ap
43	27	2.7	54	37	US-09-170-373-3889	Sequence 3889, Ap
44	22	2.2	123	28	US-10-221-279-9377	Sequence 9377, Ap
45	22	2.2	669	27	US-10-170-2058-22195	Sequence 22195, A

ALIGNMENTS

RESULT 1  
US-09-914-042-1  
; Sequence 1, Application US/09914042  
; GENERAL INFORMATION:  
; APPLICANT: SCHLESSINGER, JOSEPH  
; APPLICANT: ANDREEV, JULIAN  
; TITLE OF INVENTION: PYK2 BINDING PROTEINS  
; FILE REFERENCE: 038602/1224  
; CURRENT APPLICATION NUMBER: US/09/914.042  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: PCT/US00/04647  
; PRIOR FILING DATE: 2000-02-18  
; PRIOR APPLICATION NUMBER: 60/121,125  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1006  
; TYPE: PRT

i ORGANISM: Homo sapiens  
US-09-914-042-1

Query Match 100.0%; Score 1006; DB 24; Length 1006;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA	60
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA	60
Qy	61	INSSGLAHVNEEQYTOALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120
Db	61	INSSGLAHVNEEQYTOALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120
Qy	121	NIISFPLDSLKGLDKGVKGDLPKDFKAWDKYETKITKIEKKEHAKLHGMIRTEISG	180
Db	121	NIISFPLDSLKGLDKGVKGDLPKDFKAWDKYETKITKIEKKEHAKLHGMIRTEISG	180
Qy	181	AEIAEEMEKERRFFOLQWCEYLLKVNELKIKKGVLLQNLKIKYHQAQCNFPQDGLKAVES	240
Db	181	AEIAEEMEKERRFFOLQWCEYLLKVNELKIKKGVLLQNLKIKYHQAQCNFPQDGLKAVES	240
Qy	241	LKPSIETLSTDLHTTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQIRQSTAYSILHQPG	300
Db	241	LKPSIETLSTDLHTTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQIRQSTAYSILHQPG	300
Qy	301	NKEHGTERRNGSLYKKSQDGIKRWQKRCVKNQGFITISHGTANRPPAKNLITCQVKTNP	360
Db	301	NKEHGTERRNGSLYKKSQDGIKRWQKRCVKNQGFITISHGTANRPPAKNLITCQVKTNP	360
Qy	361	KEKCFDILSHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT	420
Db	361	KEKCFDILSHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT	420
Qy	421	KEIISEVQRMGTNDVCCDGPDPDWLSTNLGILTCIECSGIHRELGVHSPMQSLTLDV	480
Db	421	KEIISEVQRMGTNDVCCDGPDPDWLSTNLGILTCIECSGIHRELGVHSPMQSLTLDV	480
Qy	481	LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNFGSDMNARKDYITAKYIERRYARKKH	540
Db	481	LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNFGSDMNARKDYITAKYIERRYARKKH	540
Qy	541	ADNAKLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS	600
Db	541	ADNAKLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS	600
Qy	601	PLTNKGQPRGPDVLSATLALGPLSNAMVLQPPAPMPKRSQATKLKPKVKALYNCAVNP	960
Db	901	PLTNKGQPRGPDVLSATLALGPLSNAMVLQPPAPMPKRSQATKLKPKVKALYNCAVNP	960
Qy	961	DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVPSFVHFIAAD 1006	

Db 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVPSFVHFIAAD 1006

RESULT 2

US-09-791-537-142678  
; Sequence 142678, Application US/09791537  
; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 142678

; LENGTH: 1006

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-791-537-142678

Query Match 69.9%; Score 703; DB 22; Length 1006;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA	60
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA	60
Qy	61	INSSGLAHVNEEQYTOALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120
Db	61	INSSGLAHVNEEQYTOALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120
Qy	121	NIISFPLDSLKGLDKGVKGDLPKDFKAWDKYETKITKIEKKEHAKLHGMIRTEISG	180
Db	121	NIISFPLDSLKGLDKGVKGDLPKDFKAWDKYETKITKIEKKEHAKLHGMIRTEISG	180
Qy	181	AEIAEEMEKERRFPQLOMCEYLLKVNELKIKKGVLLQNLKIKYHQAQCNFPQDGLKAVES	240
Db	181	AEIAEEMEKERRFPQLOMCEYLLKVNELKIKKGVLLQNLKIKYHQAQCNFPQDGLKAVES	240
Qy	241	LKPSIETLSTDLHTTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQIRQSTAYSILHQPG	300
Db	241	LKPSIETLSTDLHTTIKQAQDEERRQLIQRLDILKSALQVEQKEDSQIRQSTAYSILHQPG	300
Qy	301	NKEHGTERRNGSLYKKSQDGIKRWQKRCVKNQGFITISHGTANRPPAKNLITCQVKTNP	360
Db	301	NKEHGTERRNGSLYKKSQDGIKRWQKRCVKNQGFITISHGTANRPPAKNLITCQVKTNP	360
Qy	361	BEKCFDILSHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT	420
Db	361	BEKCFDILSHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT	420
Qy	421	KEIISEVQRMGTNDVCCDGPDPDWLSTNLGILTCIECSGIHRELGVHSPMQSLTLDV	480
Db	421	KEIISEVQRMGTNDVCCDGPDPDWLSTNLGILTCIECSGIHRELGVHSPMQSLTLDV	480
Qy	481	LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNFGSDMNARKDYITAKYIERRYARKKH	540
Db	481	LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNFGSDMNARKDYITAKYIERRYARKKH	540
Qy	541	ADNAKLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS	600
Db	541	ADNAKLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS	600
Qy	601	LHIIVDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI	660
Db	601	LHIIVDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI	660
Qy	661	AKRLKHEHCEBELLTQALSGRFSNHHVEYEWRLLEDHEDLDESDDMDKELQSPENRRDRP	720

Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHLEDDESDDDMDEKLPSPNRREDRP 720  
Qy 721 ISFYQLGSNQLQSNVSLARDAANLAKKQAFMPSILQNETYCALLSGSPPPAQAAPS 780  
Db 721 ISFYQLGSNQLQSNVSLARDAANLAKKQAFMPSILQNETYCALLSGSPPPAQAAPS 780  
Qy 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPPLPLRVSTN 840  
Db 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPPLPLRVSTN 840  
Qy 841 PLTPTPPPPVAKTSPVMEALSOQSKPAPPGISQIRPPPLPQPPSRLPQKPPACTDKST 900  
Db 841 PLTPTPPPPVAKTSPVMEALSOQSKPAPPGISQIRPPPLPQPPSRLPQKPPACTDKST 900  
Qy 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVADNP 960  
Db 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVADNP 960  
Qy 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFVFIAD 1006  
Db 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFVFIAD 1006

RESULT 3  
US-09-949-002-361  
; Sequence 361, Application US/09949002  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION  
; FILE REFERENCE: AND USES THEREOF  
; FILE REFERENCE: CL000790  
; CURRENT APPLICATION NUMBER: US/09/949,002  
; CURRENT FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/231,401  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 10823  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 361  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-002-361

Query Match 69.9%; Score 703; DB 24; Length 1006;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60  
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60

Qy 61 INSSGLAHVENEEOYTOALEKFGGNCVCRDDPDILGSAPLKFESVTKELTALFKNLQNMN 120  
Db 61 INSSGLAHVENEEOYTOALEKFGGNCVCRDDPDILGSAPLKFESVTKELTALFKNLQNMN 120

Qy 121 NIISFPDLSLKGDLKGVGDLKPPDKAWKDYETKITKEKEKEHAKLHGMRTISG 180  
Db 121 NIISFPDLSLKGDLKGVGDLKPPDKAWKDYETKITKEKEKEHAKLHGMRTISG 180

Qy 181 AEIAEEMEKERRFFQLOMCEYLLKNEIKIKKGVDDLLQNLKIKYFHAQCNFFQDLKAVES 240  
Db 181 AEIAEEMEKERRFFQLOMCEYLLKNEIKIKKGVDDLLQNLKIKYFHAQCNFFQDLKAVES 240

Qy 241 LKPSIETLSTDLHTIKQAQDERQLIOLRILKSALQVEQKDSQIRQSTAYSILHQPOG 300  
Db 241 LKPSIETLSTDLHTIKQAQDERQLIOLRILKSALQVEQKDSQIRQSTAYSILHQPOG 300

Qy 301 NKEHGTERRGSLYKKSQDGRKVKWQKCSVXNGFLTISHGTANRPPAKNLNLTQVKTNP 360  
Db 301 NKEHGTERRGSLYKKSQDGRKVKWQKCSVXNGFLTISHGTANRPPAKNLNLTQVKTNP 360

Qy 361 EEKCFDLISHDRTYHFOAEDEQECQIWMVSLQNSKEBALNNAFKGDDNTGNNIVQBLT 420

Db 361 EEKCFDLISHDRTYHFOAEDEQECQIWMVSLQNSKEBALNNAFKGDDNTGNNIVQBLT 420  
Qy 421 KEIISVQVMTGNDVCCDCCAPDPTWLSTNLGILTCIECSGIHRELGVHVPMSQSLTLDV 480  
Db 421 KEIISVQVMTGNDVCCDCCAPDPTWLSTNLGILTCIECSGIHRELGVHVPMSQSLTLDV 480  
Qy 481 LGTSELALLAKNIGNAGFNEIMECCLPABEDSVKPNPGSDMMARKDYITAKYIERYARKKH 540  
Db 481 LGTSELALLAKNIGNAGFNEIMECCLPABEDSVKPNPGSDMMARKDYITAKYIERYARKKH 540  
Qy 541 ADNAKHLHSLCEAVKTRDIFGLLOAYADGVLDLTKIPLANGHEDEETALHLAVRSVDRTS 600  
Db 541 ADNAKHLHSLCEAVKTRDIFGLLOAYADGVLDLTKIPLANGHEDEETALHLAVRSVDRTS 600  
Qy 601 LHIIVDFLVQNSGNDLKOTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660  
Db 601 LHIIVDFLVQNSGNDLKOTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660  
Qy 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHLEDDESDDDMDEKLPSPNRREDRP 720  
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHLEDDESDDDMDEKLPSPNRREDRP 720  
Qy 721 ISFYQLGSNQLQSNVSLARDAANLAKKQAFMPSILQNETYCALLSGSPPPAQAAPS 780  
Db 721 ISFYQLGSNQLQSNVSLARDAANLAKKQAFMPSILQNETYCALLSGSPPPAQAAPS 780  
Qy 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPPLPLRVSTN 840  
Db 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSGRQRSSDDPPAVHPPPLPLRVSTN 840  
Qy 841 PLTPTPPPPVAKTSPVMEALSOQSKPAPPGISQIRPPPLPQPPSRLPQKPPACTDKST 900  
Db 841 PLTPTPPPPVAKTSPVMEALSOQSKPAPPGISQIRPPPLPQPPSRLPQKPPACTDKST 900  
Qy 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVADNP 960  
Db 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVADNP 960  
Qy 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFVFIAD 1006  
Db 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFVFIAD 1006

RESULT 4  
US-09-949-003C-2272  
; Sequence 2272, Application US/09949003C  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC  
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL000791  
; CURRENT APPLICATION NUMBER: US/09/949,003C  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/231,446  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 74065  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2272  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-949-003C-2272

Query Match 69.9%; Score 703; DB 24; Length 1006;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60  
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60

Qy 61 INSSGLAHVENEEOYTOALEKFGGNCVCRDDPDILGSAPLKFESVTKELTALFKNLQNMN 120

```
|||||
Db 61 INSSGLAHVNEEQYTOALEKFGNCVCRDDPDLSAFLKESVFTKELTALFKNLIQNMN 120
Qy 121 NIISFPDLSLLKGDLLKGVKGLDKKPFDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 121 NIISFPDLSLLKGDLLKGVKGLDKKPFDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Qy 181 AEIAEEMEKERRFPQLOQCEVLLKNEIKIKKGVDLLQNLIKYFHAQCNFPQDGLKAVES 240
Db 181 AEIAEEMEKERRFPQLOQCEVLLKNEIKIKKGVDLLQNLIKYFHAQCNFPQDGLKAVES 240
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Db 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Qy 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNNGFLTISHGTANRPPAKNLNLLTCQVKTNP 360
Db 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNNGFLTISHGTANRPPAKNLNLLTCQVKTNP 360
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT 420
Db 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT 420
Qy 421 KEIISEVQRMGTNDVCCDCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMOSSLTLDV 480
Db 421 KEIISEVQRMGTNDVCCDCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMOSSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERRYARKKH 540
Qy 541 ADNAAKLHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Db 541 ADNAAKLHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Qy 601 LHIYDVLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIYDVLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Db 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Qy 721 ISFYQLGNSQLQSNNAVSLARDAANLAKKORAFMPSILQNETYGALLSGSPPPQAAPAS 780
Db 721 ISFYQLGNSQLQSNNAVSLARDAANLAKKORAFMPSILQNETYGALLSGSPPPQAAPAS 780
Qy 781 TTSAAPLPVRNVKQVQTASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Db 781 TTSAAPLPVRNVKQVQTASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Qy 841 PLTPPPPPVAKTPSVMEALSQSPKAPPGISQIRPPPLPQPPSRRLPQKPAAGTKST 900
Db 841 PLTPPPPPVAKTPSVMEALSQSPKAPPGISQIRPPPLPQPPSRRLPQKPAAGTKST 900
Qy 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCAVNP 960
Db 901 PLTNKGQPRGVDLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCAVNP 960
Qy 961 DELTSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPFVSFVHFIAID 1006
Db 961 DELTSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPFVSFVHFIAID 1006
```

## RESULT 5

```
US-09-949-002-425
; Sequence 425, Application US/09949002
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
```

```
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 425
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-425

Query Match 68.9%; Score 693; DB 24; Length 911;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MPDQISVSSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLKMKKS VKA 60
Db 17 MPDQISVSSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLKMKKS VKA 76
Qy 61 INSSGLAHVNEEQYTOALEKFGNCVCRDDPDLSAFLKESVFTKELTALFKNLIQNMN 120
Db 77 INSSGLAHVNEEQYTOALEKFGNCVCRDDPDLSAFLKESVFTKELTALFKNLIQNMN 136
Qy 121 NIISFPDLSLLKGDLLKGVKGLDKKPFDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 137 NIISFPDLSLLKGDLLKGVKGLDKKPFDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 196
Qy 181 AEIAEEMEKERRFPQLOQCEVLLKNEIKIKKGVDLLQNLIKYFHAQCNFPQDGLKAVES 240
Db 197 AEIAEEMEKERRFPQLOQCEVLLKNEIKIKKGVDLLQNLIKYFHAQCNFPQDGLKAVES 256
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Db 257 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 316
Qy 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNNGFLTISHGTANRPPAKNLNLLTCQVKTNP 360
Db 317 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNNGFLTISHGTANRPPAKNLNLLTCQVKTNP 376
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT 420
Db 377 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQELT 436
Qy 421 KEIISEVQRMGTNDVCCDCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMOSSLTLDV 480
Db 437 KEIISEVQRMGTNDVCCDCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMOSSLTLDV 496
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 497 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNARKDYITAKYIERRYARKKH 556
Qy 541 ADNAAKLHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Db 557 ADNAAKLHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 616
Qy 601 LHIYDVLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 617 LHIYDVLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 676
Qy 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Db 677 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 736
Qy 721 ISFYQLGNSQLQSNNAVSLARDAANLAKKORAFMPSILQNETYGALLSGSPPPQAAPAS 780
Db 737 ISFYQLGNSQLQSNNAVSLARDAANLAKKORAFMPSILQNETYGALLSGSPPPQAAPAS 796
Qy 781 TTSAAPLPVRNVKQVQTASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Db 797 TTSAAPLPVRNVKQVQTASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 856
Qy 841 PLTPPPPPVAKTPSVMEALSQSPKAPPGISQIRPPPLPQPPSRRLPQKPAAGTKST 895
```

Db 857 PLTPTPPPPVAKTSPVMEALSQSPKAPPAGISQIRPPPLPQPPSRLPQKKPAPG 911  
|||||  
RESULT 6  
US-09-949-003C-2770  
; Sequence 2770, Application US/09949003C  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC  
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL000791  
; CURRENT APPLICATION NUMBER: US/09/949,003C  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/231,446  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 74065  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2770  
; LENGTH: 911  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-949-003C-2770  
Query Match 68.9%; Score 693; DB 24; Length 911;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60  
|||||  
Db 17 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 76  
|||||  
Qy 61 INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGS AFLKFSVFTKELTALFKNLIQNMN 120  
|||||  
Db 77 INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGS AFLKFSVFTKELTALFKNLIQNMN 136  
|||||  
Qy 121 NIISFPLDSLKGLDKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180  
|||||  
Db 137 NIISFPLDSLKGLDKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 196  
|||||  
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFAQC NFPDQGLKAVES 240  
|||||  
Db 197 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFAQC NFPDQGLKAVES 256  
|||||  
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYS LHQPQG 300  
|||||  
Db 257 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYS LHQPQG 316  
|||||  
Qy 301 NKEHGTERNGSLYKSDGIRKWKRCVKNQGLTISHGTANRPPAKNL LITCOVKTNP 360  
|||||  
Db 317 NKEHGTERNGSLYKSDGIRKWKRCVKNQGLTISHGTANRPPAKNL LITCOVKTNP 376  
|||||  
Qy 361 BEKCFDLISHDRTYHFQAEDEQECQIWMVLSQNSKEEALNNAFKGDDNTGNNIVBELT 420  
|||||  
Db 377 BEKCFDLISHDRTYHFQAEDEQECQIWMVLSQNSKEEALNNAFKGDDNTGNNIVBELT 436  
|||||  
Qy 421 KEIISVORMTGNVCCDCGAPDPTWLSNLGILTCIECSGIHRELGVHYSRMSQSLTLDV 480  
|||||  
Db 437 KEIISVORMTGNVCCDCGAPDPTWLSNLGILTCIECSGIHRELGVHYSRMSQSLTLDV 496  
|||||  
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 540  
|||||  
Db 497 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 556  
|||||  
Qy 541 ADNAKHLISCEAVKTRDIFGLQYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600  
|||||  
Db 557 ADNAKHLISCEAVKTRDIFGLQYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 616  
|||||  
Qy 601 LHI VDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLGRKASIBIANESGETPLDI 660  
|||||  
Db 617 LHI VDFLVQNSGNDKQTKGSTALHYCCLTDNAECLLGRKASIBIANESGETPLDI 676  
|||||  
Qy 661 AKRLKHEHCEBLLTQALSGRFNSHVHVEYEWRLLEDHEDDDEKQPSNNRREDRP 720  
|||||

Db 677 AKRLKHEHCEBLLTQALSGRFNSHVHVEYEWRLLEDHEDDDEKQPSNNRREDRP 736  
|||||  
Qy 721 ISFYQLGSNQLQSNVSLARDAANLAKEKQRAFMPSTILQNETY GALLSGSPPPAQPAAPS 780  
|||||  
Db 737 ISFYQLGSNQLQSNVSLARDAANLAKEKQRAFMPSTILQNETY GALLSGSPPPAQPAAPS 796  
|||||  
Qy 781 TTSAPPLPPRVNGVKQVTASSANTLWKTNSVSDGSGRQRSSSDPPAVHPPLPPLRVVTSTN 840  
|||||  
Db 797 TTSAPPLPPRVNGVKQVTASSANTLWKTNSVSDGSGRQRSSSDPPAVHPPLPPLRVVTSTN 856  
|||||  
Qy 841 PLTTTPPPVAKTSPVMEALSQSPKAPPAGISQIRPPPLPQPPSRLPQKKPAPG 895  
|||||  
Db 857 PLTTTPPPVAKTSPVMEALSQSPKAPPAGISQIRPPPLPQPPSRLPQKKPAPG 911  
|||||  
RESULT 7  
US-10-170-205E-7820  
; Sequence 7820, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CL001381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7820  
; LENGTH: 990  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-170-205E-7820  
Query Match 68.9%; Score 693; DB 27; Length 990;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60  
|||||  
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60  
|||||  
Qy 61 INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGS AFLKFSVFTKELTALFKNLIQNMN 120  
|||||  
Db 61 INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGS AFLKFSVFTKELTALFKNLIQNMN 120  
|||||  
Qy 121 NIISFPLDSLKGLDKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180  
|||||  
Db 121 NIISFPLDSLKGLDKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180  
|||||  
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFAQC NFPDQGLKAVES 240  
|||||  
Db 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFAQC NFPDQGLKAVES 240  
|||||  
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYS LHQPQG 300  
|||||  
Db 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYS LHQPQG 300  
|||||  
Qy 301 NKEHGTERNGSLYKSDGIRKWKRCVKNQGLTISHGTANRPPAKNL LITCOVKTNP 360  
|||||  
Db 301 NKEHGTERNGSLYKSDGIRKWKRCVKNQGLTISHGTANRPPAKNL LITCOVKTNP 360  
|||||  
Qy 361 BEKCFDLISHDRTYHFQAEDEQECQIWMVLSQNSKEEALNNAFKGDDNTGNNIVBELT 420  
|||||  
Db 361 BEKCFDLISHDRTYHFQAEDEQECQIWMVLSQNSKEEALNNAFKGDDNTGNNIVBELT 420  
|||||  
Qy 421 KEIISVORMTGNVCCDCGAPDPTWLSNLGILTCIECSGIHRELGVHYSRMSQSLTLDV 480  
|||||  
Db 421 KEIISVORMTGNVCCDCGAPDPTWLSNLGILTCIECSGIHRELGVHYSRMSQSLTLDV 480  
|||||  
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 540  
|||||  
Db 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 540  
|||||

```

QY 541 ADNAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDRS 600
Db 541 ADNAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDRS 600
QY 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
QY 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLOPSENRRDRP 720
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLOPSENRRDRP 720
QY 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
Db 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
QY 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
QY 841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 895
Db 841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 895

RESULT 8
US-60-452-680-18115
; Sequence 18115, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18115
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-18115

Query Match 68.9%; Score 693; DB 37; Length 990;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEALDVRMVLKMKKSVA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEALDVRMVLKMKKSVA 60
QY 61 INSSGLAHVENEEOYTOALEKFGNCVCRDPPDLGSAFLKFSVFTKELTALFKNLQNMN 120
Db 61 INSSGLAHVENEEOYTOALEKFGNCVCRDPPDLGSAFLKFSVFTKELTALFKNLQNMN 120
QY 121 NIISFPLDSLKGDUGVKGDLKKPFDKAWXDYETKITIKIEKKEHAKLHGMIRTEISG 180
Db 121 NIISFPLDSLKGDUGVKGDLKKPFDKAWXDYETKITIKIEKKEHAKLHGMIRTEISG 180
QY 181 AEIAEEMEKERRFFQLOQCEVLLKYNELIKKGVLDLQNLKYPHACNFFQDGLKAVES 240
Db 181 AEIAEEMEKERRFFQLOQCEVLLKYNELIKKGVLDLQNLKYPHACNFFQDGLKAVES 240
QY 241 LKPSITETLSTDLHTTKQAQDEBERRQLIQRLDILKSALQVEQKESQIRQSTAYSILHQPG 300
Db 241 LKPSITETLSTDLHTTKQAQDEBERRQLIQRLDILKSALQVEQKESQIRQSTAYSILHQPG 300
QY 301 NKEHGTENSGLYKKSQDGIRKQWKQKCSVKNQFLTISHGTANRPPAKNLLTTCQVKNP 360
Db 301 NKEHGTENSGLYKKSQDGIRKQWKQKCSVKNQFLTISHGTANRPPAKNLLTTCQVKNP 360

```

```

QY 361 BEKKCFDLISHDRTHYFQAEDQEBCQIWMSVLQNSKEEALNNAFKGDDNTGNNIVOELT 420
Db 361 BEKKCFDLISHDRTHYFQAEDQEBCQIWMSVLQNSKEEALNNAFKGDDNTGNNIVOELT 420
QY 421 KEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSIHRELGVHYSMPQSLTLDV 480
Db 421 KEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSIHRELGVHYSMPQSLTLDV 480
QY 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMMARKDYITAKYIERYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMMARKDYITAKYIERYARKKH 540
QY 541 ADNAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDRS 600
Db 541 ADNAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHVAHSVDRS 600
QY 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
QY 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLOPSENRRDRP 720
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLOPSENRRDRP 720
QY 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
Db 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
QY 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
QY 841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 895
Db 841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 895

RESULT 9
US-60-452-680-18114
; Sequence 18114, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18114
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-18114

Query Match 26.8%; Score 270; DB 37; Length 909;
Best Local Similarity 99.7%; Pred. No. 4e-257;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 342 ANRPPAKNLLTTCQVKTNPKEKCFDLISHDRTHYFQAEDQEBCQIWMSVLQNSKEEALN 401
Db 306 ANRPPAKNLLTTCQVKTNPKEKCFDLISHDRTHYFQAEDQEBCQIWMSVLQNSKEEALN 365
QY 402 NAFKGGDDNTGNNIVQBELTKEIIESEVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSG 461
Db 366 NAFKGGDDNTGNNIVQBELTKEIIESEVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSG 425
QY 462 IHRELGVHYSMPQSLTLDVLGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNA 521
Db 426 IHRELGVHYSMPQSLTLDVLGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNA 485
QY 522 RKDYITAKYIERYARKKHADNAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANG 581

```



Db 486 RKDYITAKYIERRYARKKHAANAALSLCSAVKTRDIFGLQAYADGVDTLTKIPLANG 545  
QY 582 HEPDHALHLAVRSVDRSLHIVDFLVQNSGNLDKQTKGSGTALHYCCLTDAECLKLL 641  
Db 546 HEPDHALHLAVRSVDRSLHIVDFLVQNSGNLDKQTKGSGTALHYCCLTDAECLKLL 605  
QY 642 RGKASIEIANSGETPLDIAKRLKHEHCEELLTQALSGRFSNHHVVEYEWRLLEDDES 701  
Db 606 RGKASIEIANSGETPLDIAKRLKHEHCEELLTQALSGRFSNHHVVEYEWRLLEDDES 665  
QY 702 DDDWDEKLQPS 712  
Db 666 DDDWDEKLQPS 676

## RESULT 10

US-09-914-042-2

; Sequence 2, Application US/09914042  
; GENERAL INFORMATION:  
; APPLICANT: SCHLESINGER, JOSEPH  
; APPLICANT: ANDREEV, JULIAN  
; TITLE OF INVENTION: PVK2 BINDING PROTEINS  
; FILE REFERENCE: 038602/1224  
; CURRENT APPLICATION NUMBER: US/09/914,042  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: PCT/US00/04647  
; PRIOR FILING DATE: 2000-02-18  
; PRIOR APPLICATION NUMBER: 60/121,125  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-914-042-2

Query Match 19.4%; Score 195; DB 24; Length 826;

Best Local Similarity 100.0%; Pred. No. 7.7e-183;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 LYKSDGTRKWQKRCVKGNGFLTISHGTANRPAPKUNLITCQVKNTPBKKCFDLISH 371  
Db 140 LYKSDGTRKWQKRCVKGNGFLTISHGTANRPAPKUNLITCQVKNTPBKKCFDLISH 199  
QY 372 DRTYHFOADEQECCQIWMVQLNSKEEALNNAFKGDDNTGNNIVQELTKIISEVQRM 431  
Db 200 DRTYHFOADEQECCQIWMVQLNSKEEALNNAFKGDDNTGNNIVQELTKIISEVQRM 259  
QY 432 GNDVCCDCGADPTWLSTNLGILTCIECSGIHRELGVHYSQMOSLTLDVLGTSSELLAKN 491  
Db 260 GNDVCCDCGADPTWLSTNLGILTCIECSGIHRELGVHYSQMOSLTLDVLGTSSELLAKN 319  
QY 492 IGNAGFNEIMCCLP 506  
Db 320 IGNAGFNEIMCCLP 334

## RESULT 11

US-09-791-537-31632

; Sequence 31632, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31632  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: pdb 1DCQA  
US-09-791-537-31632

Query Match 15.0%; Score 151; DB 22; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.1e-139;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 AAKLHSLCEAVKTRDIFGLQAYADGVDTLTKIPLANGHEPDETALHLAVRSVDRSLHI 603  
Db 128 AAKLHSLCEAVKTRDIFGLQAYADGVDTLTKIPLANGHEPDETALHLAVRSVDRSLHI 187  
QY 604 VDFLVQNSGNLDKQTKGSGTALHYCCLTDAECLKLLRGKASIEIANSGETPLDIAKR 663  
Db 188 VDFLVQNSGNLDKQTKGSGTALHYCCLTDAECLKLLRGKASIEIANSGETPLDIAKR 247  
QY 664 LKHEHCEELLTQALSGRFSNHHVVEYEWRL 694  
Db 248 LKHEHCEELLTQALSGRFSNHHVVEYEWRL 278

## RESULT 12

US-10-104-047-2284

; Sequence 2284, Application US/10104047  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2284  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2284

Query Match 10.9%; Score 110; DB 27; Length 208;

Best Local Similarity 100.0%; Pred. No. 3.4e-99;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 897 DKSTPLTNKGOPRGVPLSATEALGPLSNAMVLQPPAMPBKSQATKLKPKRVKALYN 956  
Db 99 DKSTPLTNKGOPRGVPLSATEALGPLSNAMVLQPPAMPBKSQATKLKPKRVKALYN 158  
QY 957 ADNPDELTFSEGDIIVDGEDQEWIGHIDGDPGRKGAFVSVFHFAD 1006  
Db 159 ADNPDELTFSEGDIIVDGEDQEWIGHIDGDPGRKGAFVSVFHFAD 208

## RESULT 13

PCT-US01-01332-635

; Sequence 635, Application PC/TUS0101332  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc., et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PUZ02PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/01332  
; CURRENT FILING DATE: 2001-05-09  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 635  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE

; LOCATION: (191)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-01332-635

Query Match 10.9%; Score 110; DB 1; Length 350;  
Best Local Similarity 100.0%; Pred. No. 5.7e-99;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 897 DKSTPLTNKGPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKPKRVKALYNVCV 956  
Db 241 DKSTPLTNKGPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKPKRVKALYNVCV 300  
Qy 957 ADNPDDELTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFPVSVFVHFIAD 1006  
Db 301 ADNPDDELTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFPVSVFVHFIAD 350

## RESULT 14

US-09-764-875-635  
; Sequence 635, Application US/09764875  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ202  
; CURRENT APPLICATION NUMBER: US/09/764,875  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 635  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (191)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-875-635

Query Match 10.9%; Score 110; DB 22; Length 350;  
Best Local Similarity 100.0%; Pred. No. 5.7e-99;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 897 DKSTPLTNKGPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKPKRVKALYNVCV 956  
Db 241 DKSTPLTNKGPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKPKRVKALYNVCV 300  
Qy 957 ADNPDDELTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFPVSVFVHFIAD 1006  
Db 301 ADNPDDELTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFPVSVFVHFIAD 350

## RESULT 15

PCT-US01-01332-956  
; Sequence 956, Application PC/TUS0101332  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc., et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ202PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/01332  
; CURRENT FILING DATE: 2001-05-09  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 956  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (28)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE

; LOCATION: (65)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (66)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (178)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (190)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (270)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-01332-956

Query Match 7.9%; Score 79; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 3e-68;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 928 VLQPPAPMPRKSOATKPKRVKALYNVCADNPDELTFSEGDVIVDGEEDQEWIIGHID 987  
Db 271 VLQPPAPMPRKSOATKPKRVKALYNVCADNPDELTFSEGDVIVDGEEDQEWIIGHID 330  
Qy 988 GDCGRKGAFPVSVFVHFIAD 1006  
Db 331 GDCGRKGAFPVSVFVHFIAD 349

Search completed: August 5, 2005, 04:10:01  
Job time : 241 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:32:11 ; Search time 7497 Seconds  
(without alignments)  
5107.730 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAFVPVSVHFIAD 1006

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US0914042/runat\_04082005\_090632\_2122/app\_query.fasta\_1.1159  
-DB=EST -QFMT=fastap -SURFIX=first -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0914042 @CEN 1 1 4326 @runat\_04082005\_090632\_2122 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1592.5	30.2	1167	5	BUS08193
2	1409.5	26.7	884	5	BQ439046
3	1391	26.4	824	7	CNS35522
4	1352.5	25.6	909	5	BUI159073
5	1351	25.6	773	7	CN460643
6	1351	25.6	781	7	CO429713
7	1271	24.1	731	7	CNS38273
8	1271	24.1	742	7	CNS525450
9	1269	24.1	901	5	BUI172807

10	1266	24.0	784	6	CA328666	UI-M-FY0-
11	1260	23.9	742	7	CNS533008	UI-M-H00-
12	1246.5	23.6	740	5	BM944440	UI-M-EH0P
13	1237	23.5	738	7	CNS37253	UI-M-HS0-
14	1235	23.4	737	7	CF539222	UI-M-GH0-
15	1229	23.3	861	5	BUI22224	603488266
16	1228.5	23.3	846	5	BUI55022	AGENCOURT
17	1226	23.3	761	7	CF535252	UI-M-GH0-
18	1215	23.0	689	5	BUI703510	UI-M-F00-
19	1211	23.0	977	5	BQ882484	AGENCOURT
20	1196.5	22.7	753	7	CF535194	UI-M-GH0-
21	1196	22.7	742	7	CNS535077	UI-M-HS0-
22	1188	22.5	702	7	CF532249	UI-M-GH0-
23	1177.5	22.3	1014	4	BG293568	BUS293568
24	1174	22.3	891	5	BUI10830	AGENCOURT
25	1171	22.2	735	6	CB246523	UI-M-F10-
26	1155	21.9	714	7	CF745192	UI-M-GV0-
27	1143.5	21.7	801	5	EX914137	EX914137
28	1140	21.6	680	6	BY720163	BY720163
29	1132.5	21.5	682	4	B1562457	B1562457
30	1129.5	21.4	780	5	EX914135	EX914135
31	1103	20.9	663	7	CR769039	CR769039
32	1092.5	20.7	684	2	BE617384	601442018
33	1091	20.7	913	5	BQ422593	AGENCOURT
34	1089.5	20.7	948	6	CD358314	AGENCOURT
35	1083.5	20.5	3115	3	BC038615	Mus muscu
36	1062	20.1	629	7	CN165820	996655 MA
37	1041	19.7	702	6	CD350751	UI-M-G10-
38	1039.5	19.7	924	5	BUI253672	603746385
39	1031	19.6	612	7	CF539480	UI-M-GH0-
40	1031	19.6	741	6	CD494693	CD412-C03
41	1008.5	19.1	775	5	BUI32648	603120421
42	995.5	18.9	978	4	BM471456	AGENCOURT
43	982	18.6	795	7	CR435039	CR435039
44	977	18.5	567	7	CF735384	UI-M-HB0-
45	968	18.4	965	5	BUI51927	AGENCOURT

#### ALIGNMENTS

RESULT 1  
BUS08193  
LOCUS BUS08193 1167 bp mRNA linear EST 12-SEP-2002  
DEFINITION AGENCOURT\_10128476 NIH\_MGC\_71 Homo sapiens CDNA clone IMAGE:6502231  
5', mRNA sequence.  
ACCESSION BUS08193  
VERSION BUS08193.1 GI:22814426  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1167)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L14M14057 row: p column: 08  
High quality sequence stop: 762.  
Location/Qualifiers  
1. 1167  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6502231"

```

/tissue type="leiomyosarcoma"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/notice="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      2,986-104      Length:      1167
Score:          1592.50      Matches:      328
Percent Similarity: 77.73%      Conservative: 14
Best Local Similarity: 74.55%      Mismatches: 38
Query Match:      30.20%      Indels:      60
DB:              5          Gaps:      6

```

US-09-914-042-1 (1-1006) x BU508193 (1-1167)

```

Qy 404 PheLysGlyAspAspAsnThr-GlyGluAsnAsnIleValGlnGlnLeuThrLysGluI 423
Db 2 TTTAAGGGGATGACAAATCTGGGAGAAATTAACATGTCAGAACTGACAAAGGAGAT 61

Qy 423 eILserGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAs 443
Db 62 CATCTCAGAAGTGCAGAGATGACGGGCAATGACGCTCTGTGACTGTGGGGGCCAGA 121

Qy 443 pProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHi 463
Db 122 TCCTACATGGCTTTCACCAACCTGGGCATCCTGACCTGCATCGAGTGTTCGGGAATCCA 181

Qy 463 sArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyTh 483
Db 182 CCGAGAGCTGGGGTTCATTAATTCAGAGATGCAGTCCCTGACCTTAGATGACTGGGAAC 241

Qy 483 rSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCy 503
Db 242 ATCTGAGCTGTGCTCGCCAGAAATATTGGGAATGCAGGCTTTAATGAGATCATGGAATG 301

Qy 503 sCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLys 523
Db 302 TTGGCTACAGCTGAGGACTCAGTCCAAACCCCAACCCAGGCGGACATGAATGCAAGAAA 361

Qy 523 sAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHisAlaAspAs 543
Db 362 CGACTACATCAGCCCAAGTATCATCGAGAGGATACCGAAGGAAGACGACCGGATRA 421

Qy 543 nAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuLe 563
Db 422 CGCGCGAAGCTTTCACAGTCTTTGGGAGCGGTCAAAACGAGAGATATTTTGGATTGCT 481

Qy 563 uGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGl 583
Db 482 CCAAGCTTATGCTGATGTGTGGATCTTACGGAAAAAATCCCACTGGCCCAACGGACATGA 541

Qy 583 uProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLeuHisIl 603
Db 542 GCCGGATGAACGGCCCTCCACCTTGCAGTCAGATCCGTTGGATCGAATCTCTTTCAT 601

Qy 603 eValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerTh 623
Db 602 TGTAGACTTTTATGTTTACAGACAGTGGGAACCTGGATAAACAGACAGGGAAGGCAGCAC 661

Qy 623 rAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGl 643
Db 662 AGCCCTGCACACTCTGCTGCCTGACCGCAATGTCGAGTGCCTCAAGTTGCTCTGCGGG 721

Qy 643 YLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysAr 663
Db 722 GAAGGCTTCCATCGAGATAGCAACAGTCCAGAGAGACTCCGCTGGACATTCGCCAAGCG 781

Qy 663 gLeuLysHisGluHisCysGluGluLeuThrGlnAlaLeuSerGlyArgPheAsnSe 683
Db 782 CCTCAAGCACGAGCACTGTGAGGAGCTGTGACCCCGAGCCTTATCTGGAAAAATTTAAATTC 841

```

```

Qy 683 rHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAs 703
Db 842 TCAGTTCCTCCCTGGATATTAATGCGACTACTCCAGAAAAACCGGATGAAGTGTATGA 901

Qy 703 pAspMetAspGlu-LysLeuGlnProSerGlu-AsnArgArg---GluAspArgProIle 721
Db 902 CGACTCGATGAAAAAATTGACGCCAGTTCCTCAACCGGGGGAACACCGGCCCTTC 961

Qy 722 SerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAsp 741
Db 962 GCCTTCTACCGCGTGGG----- 979

Qy 742 AlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnGlu 761
Db 979 ----- 979

Qy 762 ThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAla----- 777
Db 980 -----CCTCCAACCAACTTTCAGCTAAACCTGTAA 1012

Qy 778 -----AlaProSerThrThrSerAlaProLeuProProProArgAsnValGlyLysVal 795
Db 1013 TCTTTGGGCCCAAAA---ATTGCTGGCCAAACCTTGCCCAAGAAAAACCAAAAGGG 1069

Qy 796 GlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGly 815
Db 1070 GCTTTTTCGCCCCAGCAATCTTTGGCCAAAATTGAAACCTTAAC-----GGG 1120

Qy 816 SerArgGlnArgSerSerAspProProAlaValHisProProLeuProPro 833
Db 1121 AGCCCCCTCTGGAGAGGGCAGACCCCC-----CCTCCCGGCCCCCA 1165

RESULT 2
BQ439046
LOCUS      BQ439046
DEFINITION AGNCOURT 7761599 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020058
5' mRNA sequence.
ACCESSION BQ439046
VERSION    BQ439046.1 GI:21178122
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 884)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Straubeberg, Ph.D.
           Email: csapbs-remail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LIAM13223 row: i column: 19
           High quality sequence stop: 733.
FEATURES   Location/Qualifiers
            1..884
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:6020058"
             /tissue type="epithelioid carcinoma"
             /lab host="DH10B (phage-resistant)"
             /clone_lib="NIH_MGC_70"
             /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
             Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
             Average insert size 1.1 kb. Library constructed by Life
             Technologies."

```

## ORIGIN

## Alignment Scores:

Pred. No.: 2,99e-91 Length: 884  
Score: 1409.50 Matches: 279  
Percent Similarity: 95.55% Conservative: 0  
Best Local Similarity: 95.55% Mismatches: 7  
Query Match: 26.73% Indels: 6  
DB: 5 Gaps: 1

US-09-914-042-1 (1-1006) x BQ439046 (1-884)

Qy 318 GlytLeArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
Db 13 GGGATCCGAAAGTGTGGCAGAAAGGAAATGTTTCAGTTAAATATGGTTTCTGACCATTA 72  
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
Db 73 TCCCATGTATCCGTTAACCGGCTCTGCAAGCTCAACCTGCTTAACCTGCAGGTGAAG 132  
Qy 358 ThrAsnProGluGluLysCysPheAspLeuLeuSerHisAspArgThrTyHisPhe 377  
Db 133 ACCAACCTCAGGAGAGAGTGTTCGACCTTATTTACATGACAGAACTTACCACTTT 192  
Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
Db 193 CAAGCTGAAGATGAACAGGAATGTCAATATATGGATGTCTGTCTGCAAAATAGCAAGAA 252  
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417  
Db 253 GAAGCTTTAAACAATGCAATTTAAGGGGGATGACAACTACTGGAGAAATAAATCATGTCCTCAA 312  
Qy 418 GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
Db 313 GAACGTCAAAAGGAGATCATCTCAGAAGTCAGAGGATGACGGCAATGACGTCTGCTGT 372  
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457  
Db 373 GACTGTGGGGGCCCATGATCTTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATC 432  
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
Db 433 GAGTGTTCGGAAATCCACCGAGAGTGGGGGTTCATTTATCCAGATGACGTCCCTGACC 492  
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
Db 493 TTAGATGTACTGGGNACATCTGAGCTGTCTGCGCAGAGATATTTGGGAATGCAGGCTTT 552  
Qy 498 AsnGluIleMetGluCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
Db 553 AATGAGATCATGGAATGTTGCTTACCAGCTCAGGACTCAGTCAAAACCCCAACCCAGGCAGC 612  
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537  
Db 613 GACATGAATGCAAGAAAGAGATCATCATCAGCCAGATGATCGAGAGAGATACGCAAGG 672  
Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
Db 673 AAGAAGCAGCGGATTAACCGCGGAGCTTACAGTCTTTGCGAGCCCGCTCAAAACGAGA 732  
Qy 558 AspIlePhe-GlyLeuLeuGlnAlaTyrAlaAspGlyVal-AspLeuThrGluLysIleP 577  
Db 733 GATATTTTNGGATTGCTTCAAGCTTATGCTGATGTTGCTGATGCTTACGGGAAAAATCC 792  
Qy 577 roLeuAlaAsnGlyHisGluPro-AspGluThrAla--LeuHisLeuAlaValArgServ 596  
Db 793 CACTGGCCAAACGAGATGAGCGGGATGAACAGCGCCCTCCACCTTTCAGTCAAGATCC 852  
Qy 596 alAsp---ArgThrSerLeuHisIle 603  
Db 853 GTGGATTGGAANCCTCTCTTACATT 878

RESULT 3

## CN535522

## LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

CN535522 824 bp mRNA linear EST 29-APR-2004  
UI-M-HSO-cqg-e-24-0-UI.r1 NIH BMAP\_HSO Mus musculus CDNA clone  
IMAGE:30670487 5', mRNA sequence.  
CN535522  
CN535522.1 GI:46863678  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 824)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [rcapbs-remail.nih.gov](mailto:rcapbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousef1.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1. 824  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30670487"  
/tissue\_type="Upper Head"  
/dev\_stage="embryo 9.5 - 10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_HSO"  
/note="Organ: Upper Head; Vector: pYX-Asc; Site: 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with Not I and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CGAACTGAAT. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

## ORIGIN

## Alignment Scores:

Pred. No.: 5,81e-90 Length: 824  
Score: 1391.00 Matches: 269  
Percent Similarity: 98.55% Conservative: 2  
Best Local Similarity: 97.82% Mismatches: 4  
Query Match: 26.38% Indels: 1  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN535522 (1-824)

Qy 214 ValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAsp 233  
Db 1 GTGATTTGCTTCAGAAATCTGATCAGTACTTTTCCGCCCCAGTGCATTTTTCAGGAT 60  
Qy 234 GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis 253  
Db 61 GGATTGAAGCAGTAGAAAGCCTCAAGCCTTCCATCGAGACGCTCTCCACGGACCTCCAC 120  
Qy 254 ThrIleLysGlnAlaGlnAspGluGluArgGlnLeuIleGlnLeuArgAspIleLeu 273

```

Db      121  ACCATCAACAGCGCCAGGATGAGGAACGGCAGACGCTGATACAACTTCGAGATATTTTG 180
Qy      274  LysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGlnSerThrAlaTyr 293
Db      181  AAGTCAGCATTTGCAAGTGGAAACAGAAAGAGACTCACAACTTCGCCAAGTACAGCCTTAC 240
Qy      294  SerLeuHisGlnProGlnGlnLysGluHisGlyThrGluArgAsnGlySerLeuTyr 313
Db      241  AGCTTACATCAACTCAGGGAACAGAGGATGGAACAGAGAGGAACGGGAACCTCTAC 300
Qy      314  LysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGly 333
Db      301  AAGAAGACGATGGATCCGGAAGTGTGCGAAGAGCAAGTGTTCGTTTAAATAATGGC 360
Qy      334  PheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThr 353
Db      361  TTCTCTACCATCTCCACGCACTGCCAACCGGCCACCTGCCAAGCTCAACCTGCTAAC 420
Qy      354  CysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArg 373
Db      421  TGCACAGTGAAGACCAATCCTGAGGAGGAAGAGTGTTCGACCTCATATCATGACACGG 480
Qy      374  ThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGln 393
Db      481  ACGTACCACTTCCAAGCGGAAGACGAACAAGATGTGAGATATGATGTCTGTACTGCAG 540
Qy      394  AsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsn 413
Db      541  AACAGCAAGGAAGAGCTCTGAACACACGCCCTTTAAGGGTGTATGACAACACTGGAGAAAT 600
Qy      414  AsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsn 433
Db      601  AACATGTCTCAAGAGCTGCCACAGGAGATCATCTCGAGGTGCAGAGATGACGGGCAT 660
Qy      434  AspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIle 453
Db      661  GACGTGTCTCGCACTGTGGGGCACA - GATTTCGACGTGGCTCTCTACCAACCTGNGCATC 719
Qy      454  LeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMet 473
Db      720  CTGACTTGTCAGATGCTCTGTGGGATCCACCGGAGCTGGGGGTTCATTACTCCAGGATG 779
Qy      474  GlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeu 488
Db      780  CAGTCCCTGACATAGATGATTTGGGAACGCTGTGAGCTTCTGCTT 824

```

## RESULT 4

```

BU159073
LOCUS      BU159073
DEFINITION AGENCOURT 7938808 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6012546
5', mRNA sequence.
ACCESSION  BU159073
VERSION     BU159073.1 GI:22672983
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13203 row: p column: 19
High quality sequence stop: 636.

```

## FEATURES

## source

## Location/Qualifiers

```

1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6012546"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      3,95e-87      Length:      909
Score:          1352.50      Matches:    277
Percent Similarity: 93.07%      Conservaive: 5
Best Local Similarity: 91.42%      Mismatches: 14
Query Match:    25.65%      Indels:     7
DB:              5              Gaps:       2
US-09-914-042-1 (1-1006) x BU159073 (1-909)

Qy      385  CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeuAsnAsnAlaPhe 404
Db      1   TGTCAAAATATGATGTCTGTGCTGCAAAATAGCAAGAAGAGCTTTAAACAATGCATT 60
Qy      405  LysGlyAspAspAsnThr-GlyGluAsnAsnIleValGlnGluLeuThrLysGluIle1 424
Db      61  AAGGGGATGACATATCTGGAGGAATAATACATCGTCCAGAACTGACNAAGGAGATCAT 120
Qy      424  eSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspPr 444
Db      121  CTCAGAAAGTCAGAGGATGACGGCAATGACGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 180
Qy      444  oThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisAr 464
Db      181  TACATGGCTTTTCCACCAACCTGGGCATCCTGACCTGATCGATCGAGTGTCCGGAATCCAC 240
Qy      464  gGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSe 484
Db      241  AGAGCTGGGGTTCATTATTCCAGGATGCAGTCCCTGACCTTAGATGTACTGGGAACATC 300
Qy      484  rGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCy 504
Db      301  TGAGCTGCTGCTGCCAAGAATATTGGGAATGCAAGGCTTTAATGAGATCATGGAATGTTG 360
Qy      504  sLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAs 524
Db      361  CCTACCAGCTGAGACTCAGTCAAAACCCACAGGCGGACATGATTAATGAGATCATGGAATG 420
Qy      524  pTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAl 544
Db      421  CTACATCAGCCCAAGTACATCGAGAGGAGATACGCAAGGAGAGACGACCGGATACGC 480
Qy      544  aAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGl 564
Db      481  GGCGAAGCTTCACAGTCTTTCCGAGGCGCTCAAAACAGAGAGATATTTTGGATGCTCCA 540
Qy      564  nAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluPr 584
Db      541  AGCTTATGCTGATGTTGATCTTACGGAATAAATCCCACTGGCCCAACGACATGAGCC 600
Qy      584  oAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLeuHisIleVa 604
Db      601  GGATGAACGGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCTCTTCACATTGG 660
Qy      604  lAspPheLeuValGlnAsnSer-GlyAsnLeuAspLysGlnThrGlyLysGlySerThrA 624
Db      661  AGACTTTTAGTTTCAAGAACAGTGGGAAACCTGGGATAAACACAGACAGGAAAGGCACAG 720

```

```

QY 624 laLeuHisTyrcysLeuThrAspAenAla-GluCysLeuLysLeuLeuLeuAargly 643
Db 721 CCCTGCACTACTGCTGCTGACATCCGAGGAGGCTCAAGTCTCTCTGCGGGG 780
QY 644 Lys-AlaSerIleGluIleAlaAenGluSerGly---GluThrProLeuAapIleAlaLy 662
Db 781 AAGGCGCTCCATCGATACCAACGAGTCCGGAGAGACTCCGCTCGGACATTCGCCA 840
QY 662 sArgLeuLysHisGluHisCysGluGlu-----LeuLeuThrGlnAlaLeuSerGlyAr 680
Db 841 GCGCGCTCAAGCACCAGCAGCTGCTGTAAGAGAGCTTCCTTGACCAAGGCTTTATCTGGA 900
QY 680 g 680
Db 901 A 901

RESULT 5
CN460643
LOCUS
DEFINITION
UI-M-HB0-coy-a-05-0-UI.r1 NIH BMAP_HB0 Mus musculus cDNA clone
IMAGE:30652708 5', mRNA sequence.
ACCESSION
CN460643
VERSION
CN460643.1 GI:46466369
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 773)
NIH-MSC http://nigms.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: ccapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.wiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 773
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30652708"
/tissue_type="whole eye"
/dev_stage="embryo 12.5, 13.5, 14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP HB0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TTATGAGT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
Developing Mouse Nervous System, supported by National
Institute of Mental Health (NIMH)."
```

## ORIGIN

Alignment Scores:

Pred. No.:

4.02e-87

Length:

773

Score: 1351.00 Matches: 250  
Percent Similarity: 98.44% Conservative: 3  
Best Local Similarity: 97.28% Mismatches: 4  
Query Match: 25.62% Indels: 0  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN460643 (1-773)

```

QY 295 LeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluArgAenGlySerLeuTyrlvs 314
Db 2 TTACATCAACTCAGGGAACCAAGAGCATGGAACAGAGAGAAACGCGAAGCTCTTACAAG 61
QY 315 LysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPhe 334
Db 62 AAGAGCGATGGGATCCGGAAGTGTGGCAGAGAGAGAGTGTTCGGTTAAAGATGGCTTC 121
QY 335 LeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCys 354
Db 122 CTCACCATCTCCACAGGCACTGCCAACCGGCACCTGCCAAGCTCAACTGCTACCTGTC 181
QY 355 GlnValLysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374
Db 182 CAGGTGAAGACCAATCCTGAGGAGAGAGAGTGTTCGACCTCATATCATCATGACAGGAG 241
QY 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAen 394
Db 242 TACCACCTCCAAAGCGAAGACGAAACAAGATGTGAGATATGGATGCTCTGTACTGAGA 301
QY 395 SerLysGluGluAlaLeuAsnAlaPheLysGlyAspAspAenThrGlyGluAenAen 414
Db 302 AGCAAGGAAGAGCTCTGAAACACGCTTTAAGGGTGATGACACACTGGAGAGAAATAAC 361
QY 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAenAen 434
Db 362 ATCGTCCAGAGCTGACCAAGAGATCATCTCGAGGTGCGAGAGATGACGGCAATGAC 421
QY 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu 454
Db 422 GTGTGCTCGCACTGTGGGGCACCAGATCCGACGTGGCTCTCTACCAACTGGGCATCCTG 481
QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474
Db 482 ACTTGATCGAGTGTCTGCGGATCCACGGGAGCTGGGGGTTCATTAATCTCCAGATGAG 541
QY 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAenIleGlyAen 494
Db 542 TCCCTGACATAGATGATTGGGAACGCTGAGCTTCTGCTTGCAGAGATATCGGGAT 601
QY 495 AlaGlyPheAenGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAen 514
Db 602 GCAGGCTTTAATGAAATTTATGAGTGTTCCTACCGTCTGAGGACNCAGTCAAGCCCAAC 661
QY 515 ProGlySerAspMetAenAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArg 534
Db 662 CCAGGCGAGTGCATGATTGCAAGGAAGGACTTACATCATCAGCCAAAGTACATGAGAG 721
QY 535 TyrAlaArgLysLysHisAlaAspAenAlaLysLeuHisSerLeuCys 551
Db 722 TATGACAGGAAGAAAGCATGCCGACACCGCGCGAGGCTCCACAGCCTTTC 772
```

## RESULT 6

CO429713

LOCUS

DEFINITION

UI-M-HX0-crw-p-02-0-UI.r1 NIH BMAP\_HX0 Mus musculus cDNA clone

IMAGE:30686473 5', mRNA sequence.

ACCESSION

CO429713

VERSION

CO429713.1

GI:49676007

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 781)

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph. D.  
Email: [csapbs-remail.nih.gov](mailto:csapbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

**FEATURES**  
**SOURCE**

```

seq primer: p1x-3,
Location/Qualifiers
1. .781
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clones IMAGE:30686473"
/tissue_type="whole eye"
/dev_stages="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"
/lab_host="DHI10B (T1 phase resistant)"
/clone_lib="NIH BMAP HX0"
/notes="Organ: Eye; Vector: pYX-Asc; Site.1: EcoR I;
Sites.2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AATAATTACG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)"

```

## ORIGIN

Alignment Scores:	
Pred. No.:	4.08e-87
Score:	1351.00
Percent Similarity:	98.08%
Best Local Similarity:	96.54%
Query Match:	25.62%
DB:	7
	0
	0
Length:	781
Matches:	251
Conservative:	4
Mismatches:	5
Indels:	0
Gaps:	0

US-09-914-042-1 (1-1006) x CO429713 (1-781)

[illegible]



gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Alignment Scores:  
Pred. No.: 2,138-81 Length: 731  
Score: 1271.00 Matches: 239  
Percent Similarity: 99.18% Conservative: 2  
Best Local Similarity: 98.33% Mismatches: 2  
Query Match: 24.10% Indels: 0  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN538273 (1-731)

Qy 262 GluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGln 281  
Db 2 GAACGGGACGACGCTGATCAACTTCGATTATTTGAGTCAGCATTCGAGTGGAAACG 61  
Qy 282 LysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsn 301  
Db 62 AAAGAGGACTCAACACTCGCCAAAGTACAGCTACAGCTTACATCACTCAGGGAAAC 121  
Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLys 321  
Db 122 AAGGAGCATGGAACAGAGAGGAAACGGAACCTCTCAAGAGAGCGATCGGATAA 181  
Qy 322 ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341  
Db 182 GGTGCGGAGAGAGAGAGTGTTCCTGTTAAATGCTTCCTCACCATCTCCACGGCACT 241  
Qy 342 AlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGlu 361  
Db 242 GCCAACCGGCCACCTGCCAAGCTCAACCTGCTAACTGCCAGTGAAGCAATCTCTGAG 301  
Qy 362 GluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAsp 381  
Db 302 CAGAGAAGAGTGTTCGACCTCATATCATAGCAGGAGCTACCACTTCCAGCGGAAGAC 361  
Qy 382 GluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsn 401  
Db 362 GAACAGGAATGTCAGATATGGATGTCTGTACTGCGAAGACAGCAGGAAGAAGCTCTGAAC 421  
Qy 402 AsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys 421  
Db 422 AACGCTTTAAGGGTGATGACAACTGGAGAAATATACATGCTCCAGAGCTGACCAAG 481  
Qy 422 GluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGlyAla 441  
Db 482 GAGATCATCTCGAGGTGCAGAGGATGACGGCAATGACGTGTCTGCGACTGTGGGCA 541  
Qy 442 ProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGly 461  
Db 542 CCAGATCCGACGTGTCTCTACCAACCTGGGCATCTCTGACTGTGATCGAGTCTCTGGG 601  
Qy 462 IleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeu 481  
Db 602 ATTACCGGGAGCTGGGGTTTCATTACTCCAGATGCGATGCTCCTGACATTAGATGATTG 661  
Qy 482 GlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMet 501  
Db 662 GGAAGCTCTGAGCTTCTGCTCCCAAGAAATATCGGAATGCAAGCTCTTAAATGAATATG 721  
Qy 502 GluCysCys 504  
Db 722 GAGTGTTC 730

RESULT 8  
CN525450  
LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

Location/Qualifiers  
1..742  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30638251"  
/tissue\_type="Upper Head"  
/dev\_stages="9.5-10.5 dpc"  
/lab\_host="DHI0B (T1 phage resistant)"  
/clone\_lib="NIH BMAP HN0"  
/note="Torgan: Head: Vector: pX-Asc: Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Alignment Scores:  
Pred. No.: 2,138-81 Length: 742  
Score: 1271.00 Matches: 242  
Percent Similarity: 99.59% Conservative: 3  
Best Local Similarity: 98.37% Mismatches: 1  
Query Match: 24.10% Indels: 0  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN525450 (1-742)

Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGlyAlaGluIle 183  
Db 3 AAAAGGAGCATGCCAGCTCCAGGGATGATTCGTACTGAAATAAGGGGCTGAGATT 62  
Qy 184 AlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeu 203  
Db 63 GCAGAGGAGATGGAGAGGAAAGACGGTTCCTCCAGCTGCAGATGTGTGAGTACCTGCTG 122

CN525450 742 bp mRNA linear EST 29-APR-2004  
UI-M-HN0-cnff-f-20-0-UI.r1 NIH\_BMAP\_HN0 Mus musculus cDNA clone  
IMAGE:30638251 5', mRNA sequence.

CN525450  
CN525450.1 GI:46853103

EST.  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 742)  
NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mouseefi.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pX-5.

Location/Qualifiers

1..742

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:30638251"

/tissue\_type="Upper Head"

/dev\_stages="9.5-10.5 dpc"

/lab\_host="DHI0B (T1 phage resistant)"

/clone\_lib="NIH BMAP HN0"

/note="Torgan: Head: Vector: pX-Asc: Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

```

204 LysValAsnGluLeuLysGlyValAspLeuLeuGlnAsnLeuLeuLysTyr 223
Db 123 AAGGTCATGAATCAAGTGAAAGAGGAGTGGATTGCTTCAGATCTGATCAAGTAC 182
Qy 224 PheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243
Db 183 TTTCAAGCCAGTGCATTTTTCAGGATGGATTGAAAGCAGTAGAAGCCTCAAGCCT 242
Qy 244 SerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluArg 263
Db 243 TCCATCGAGAGCGCTCCACGAGACCTCCACACCATCAACAGAGCCAGGATGAGGAACGG 302
Qy 264 ArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGlu 283
Db 303 CGACAGCTGATACAACCTTCGAGATATTTGAAGTCAGCATTCGAGTGGAAACAGAAAGAG 362
Qy 284 AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGlu 303
Db 363 GACTCACAACTTCGCCAAAGTACAGCCTACAGCTTACATCAACCTCAGGGAACAAGGAG 422
Qy 304 HisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrp 323
Db 423 CATGAAACAGAGAGGAAGCGGAACCTCTACAGAAAGAGCGATGGGATCCGAAAGTGTGG 482
Qy 324 GlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsn 343
Db 483 CAGAAGAGAGAGTGTTCGTTTAAATATGGCTTCTCACCATCTCCACGGCAGCTGCCAAC 542
Qy 344 ArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGluGluLys 363
Db 543 CGGCCACCTGCCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAATCCTGAGGAGAAG 602
Qy 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
Db 603 AAGTGTTCGACCTCATATCATACAGAGGAGTACCACTTCCAAAGCGGAAGACGAACA 662
Qy 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403
Db 663 GAATGTCAGATATGGATGCTGTACTGCAGAACACAGCAAGAAAGAAAGAGCTCTGAAACGCC 722
Qy 404 PheLysGlyAspAspAsn 409
Db 723 TTAAAGGTGATGACAAC 740

RESULT 9
LOCUS BUI72807
DEFINITION AGENCOURT_7966859_NIH_MGC_67_Homo_sapiens_cDNA_clone_IMAGE:6170038
5', mRNA sequence.
ACCESSION BUI72807
KEYWORDS BUI72807.1 GI:22686791
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLA013536 row: b column: 23
High quality sequence stop: 609.
Location/Qualifiers
1..901

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:6170038"
/tissue_types="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."

ORIGIN
Alignment Scores: 4e-81 Length: 901
Pred. No.: 1269.00 Matches: 257
Score: 1269.00
Percent Similarity: 91.93% Conservative: 5
Best Local Similarity: 90.18% Mismatches: 16
Query Match: 24.07% Indels: 7
DB: Gaps: 2

US-09-914-042-1 (1-1006) x BUI72807 (1-901)
Qy 121 AsnIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 8 TCCGTAATCTCTCCCTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTGAAGGG 67
Qy 141 AspLeuLysLysProPheAspLeuAlaTyrLysAspTyrGluThrLysIleThrLysIle 160
Db 68 GATCTGAAAAAGCCCTTTTGTATAAAGCTTGGAAAGGACTATGAACAACAAAAATAACCAAGATA 127
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
Db 128 GAAAGAGAGAAAAAGGAACAGCCAGCTCATCGGATGATTCGGACTGGAATTAAGCGGA 187
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 188 GCGGAATTCGCGAAGAGATGGAAGAGAGAGGCGCTTCTCCAGCTACAGATGTGGAG 247
Qy 201 TyrLeuLysValAsnGluIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 248 TATCTGCTGAAGGTCAACGAAATCAAGATTAAGAGGAGTAGATTTTACTTCAGATCTG 307
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 308 ATCAAAATCTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAGAGCGTGGAAAGC 367
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 368 CTCAAACTTCATTTGAAACGCTGCTACGGATCTTCACAGCATCAACAGCCCGCAT 427
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLysSerAlaLeuGlnValGlu 280
Db 428 GAAGAAAGAGAGCAGTTGATACAGCTTCGAGATATTTTGAATCCGATTCGAGTTGAA 487
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 488 CAGAAAGAGGAGTCCCAAAATTCGTCAGAGCAGACCTTATAGCTTACATCAGCCTCAGGA 547
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 548 AACAAAGNACATGGGACCGGACCGGACCGGACCTCTACAAGAAAGAGTGCAGGATCGA 607
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 608 ANAGTGTGGCAGAAAAAGGAATGTTTCAGTAAAAATGGTTTTTCTGACCATATCCCATGGT 667
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 668 ACCGCTAACCGGCTCTCTGCAAAAGCTAACCTGTAACCTGCCAGGGGAAGACCCCT 727
Qy 361 GluGluLysLysCys-PheAspLeu-IleSerHisAspArg-ThrTyrHisPheGlnAla 379

```

728 GAGGAGAAGAGTCTTTTGACCTTTATTACCATGACAGAACTTACCACCTTTCAAGG 787

380 gluaSpGlu-----GlnGluCysGlnfile-----TtpMetSerValLeuGlnAsnSer 395

788 CTGGAAGAAGTCAACAGCAGGAATGTCCAAATAATGGATGGTCTGCGGCTGGCCAAATTTAG 847

396 LysGluGlu 398

848 CCCAAGGAA 856

RESULT 10

CA328666

LOCUS

DEFINITION

UI-M-FY0-cda-o-22-0-UI.r1 NIH BMAP\_FY0 Mus musculus cDNA clone

IMAGE:6827087 5', mRNA sequence.

CA328666

CA328666.1 GI:24546764

EST.

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 784)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cqapbs-remail.nih.gov](mailto:cqapbs-remail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyX-5.

Location/Qualifiers

1..784

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clones IMAGE:6827087"

/tissue\_type="whole brain"

/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab\_host="DH10B (TI phage resistant)"

/clone\_lib="NIH BMAP FY0"

/note="Organ: Brain; Vector: pyX- Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

ORIGIN

Alignment Scores:

Pred. No.:	5,39e-81	Length:	784
Score:	1265.00	Matches:	243
Percent Similarity:	96.15%	Conservative:	7
Best Local Similarity:	93.46%	Mismatches:	10
Query Match:	24.01%	Indels:	0
DB:	6	Gaps:	0

Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 742  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30657089"  
/tissue\_type="Upper Head"  
/dev\_stages="9.5-10.5 dpc"  
/lab\_host="DH10B (TI phage resistant)"  
/clone\_lib="NIH BMAP H00"  
/note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CGAATGTAAT. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

FEATURES  
source

# ORIGIN

Alignment Scores:  
Pred. No.: 1,35e-80 Length: 742  
Score: 1260.00 Matches: 243  
Percent Similarity: 99.60% Conservative: 3  
Best Local Similarity: 98.38% Mismatches: 1  
Query Match: 23.90% Indels: 1  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CMS33008 (1-742)

QY 192 ArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluLeuLysLeuLys 211  
DB 2 CGGTTCCTCCAGTCGACATGTGTGAGTACCTGCTGGAAGGTCATGAATCAAGGTGAAA 61  
QY 212 LysGlyValAspLeuLeuGlnAsnLeuLeuLysTyrPheHisAlaGlnCysAsnPhePhe 231  
DB 62 AAGGAGTGGATTTGCTTCAGAAATCTGATCAAGTACTTTTCAGCCCGCATTAATTTTTC 121  
QY 232 GlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAsp 251  
DB 122 CAGGATGGATTGAAGAGCAGTAGAAGCCTCAAGCCTTCATCGAGACCTTCCACGGAC 181  
QY 252 LeuHisThrIleLysGlnAlaGlnAspGluGluArgArgGlnLeuIleGlnLeuArgAsp 271  
DB 182 CTCACACCATCAACACAGCCCGAGGATGAGGAACGGCGACAGCTGATACAACTTCGAGAT 241  
QY 272 IleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGlnSerThr 291  
DB 242 ATTTTGAAGTCAGCATTCAGGTGGACAGAAAGAGGACTCACAACTTCGCCAAGGTACA 301  
QY 292 AlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySer 311  
DB 302 GCCTACACCTTACATCAACCTCAGGGAACACAGGAGCATGGACAGAGAGACGGGAAC 361  
QY 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331  
DB 362 CTCACAAGAGAGCGATGGGATCGGAAAGTGTGGCAGAGAGAGAGTGTTCGCTTAAA 421

QY 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeu 351  
DB 422 AATGGCTTCTTCACCATCTCCACGGCACTGCCAACCGCCACCTGCCAGCTCAACTG 481  
QY 352 LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371  
DB 482 CTAACCTGCCAGGTGAAGACCAATCTCTGAGGAGAAGAAGTGTTCGACCTCATATCAT 541  
QY 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerVal 391  
DB 542 GACAGGACGTACCACTTCCAAAGCGGAAGACCAAGAATGTTCAGATATGATGCTCTGTA 601  
QY 392 LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGly 411  
DB 602 CTGCAGAACAGCAGGAAGAAGCTCTGAACAACCCCTTTAAGGGTGTATGACCAACACTGGA 661  
QY 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleLeuSerGluValGlnArgMetThr 431  
DB 662 GAAATAC-ATCGTCCCAAGAGCTGACCAAGAGATCATCTCGGAGGTGCAGAGGATGACG 720  
QY 432 GlyAsnAspValCysCysAsp 438  
DB 721 GGCAATGACGTGTCTGCTCGAC 741

## RESULT 12

BM944240

LOCUS

DEFINITION

UI-M-EH0p-bvq-c-06-0-UI.x1 NIH\_BMAP\_EH0p Mus musculus cDNA clone

IMAGE:5695421 5', mRNA sequence.

ACCSSION

BM944240

VERSION

BM944240.1 GI:19427825

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Mus musculus

(house mouse)

REFERENCE

1 (bases 1 to 740)

NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 740

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:5695421"

/tissue\_type="whole brain"

/dev\_stages="embryo 18.5 dpc"

/lab\_host="DH10B (TI phage resistant)"

/clone\_lib="NIH\_BMAP\_EH0p"

/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag

sequence located between the Not 1 site and the polyA tail, is CAGCCAGCAG. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Alignment Scores:  
Pred. No.: 1,26e-79 Length: 740  
Score: 1246.50 Matches: 240  
Percent Similarity: 98.78% Conservatives: 3  
Best Local Similarity: 97.56% Mismatches: 0  
Query Match: 23.64% Indels: 3  
DB: 5 Gaps: 1

US-09-914-042-1 (1-1006) x BM944240 (1-740)

```
Qy 174 IleaArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGluArgPhe 193
Db 3 ATTCTACTGAATTAAGCGGGCTGAGATTGCAGAGGAGATGGAGAGAAAGACGGTTC 62
Qy 194 PheGlnLeuGlnMetCysGluTyrLeuLeuLysValAenGluIleLysIleLysGly 213
Db 63 TTCAGCTGCGATGTGTGAGTACTGCTGAAGGTCAATGAATCAAGTGAAGAAAGGA 122
Qy 214 ValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAenPheGlnAsp 233
Db 123 GTGGATTTCCTCAGAACTGATCAAGTACTTTTACGCCCGCAGTGCAATTTTTCAGGAT 182
Qy 234 GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis 253
Db 183 GGATTGAAAGCAGTAGTAAAGGCTCAAGCTTCCATCGAGAGCGCTCTCCAGCGACCTCCAC 242
Qy 254 ThrIleLysGlnAlaGlnAspGluArgGlnLeuIleGlnLeuArgAspIleLeu 273
Db 243 ACCATCAACAGCGCCAGGATGAGAAACGGGACAGCTGATACAACTTCGAGATATTTTG 302
Qy 274 LysSerAlaLeuGlnValGluGlnLysGlu-----AspSerGlnIleArgGlnSer 290
Db 303 AAGTCAGATTCAGGTGGAACAGAAAGTCTAGGAGAGACTCAACTTCGCCAAAGT 362
Qy 291 ThrAlaTyrSerLeuHisGlnProGlnGlyAenLysGluHisGlyThrGluArgAsnGly 310
Db 363 ACAGCTTACAGCTTACATCACTCAGCGAAACAGAGAGCATGGAACAGAGAGGAAACGGG 422
Qy 311 SerLeuTyrLysLysSerAspGlyIleArgLysValTropGlnLysArgLysCysSerVal 330
Db 423 AACCTCTCAAGAGAGCGGATGGGATCCGGAAGTGTGGCAGAGAGAAAGTGTTCGGTT 482
Qy 331 LysAenGlyPheLeuThrIleSerHisGlyThrAlaAenArgProProAlaLysLeuAen 350
Db 483 AAAAATGGCTTCCTCACCATCTCCACCGGAAGTCCCAACCGGCCACTTCGCNAGTCAAC 542
Qy 351 LeuLeuThrCysGlnValLysThrAenProGluGluLysLysCysPheAspLeuIleSer 370
Db 543 CTGCTAACCTGCAGGTGAAGACCAATCTCGAGGAGAAGAGTGTTCGACCTCATATCA 602
Qy 371 HisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTropMetSer 390
Db 603 CATGACAGGAGCTACCACTTCACGCGGAAGACGAACAAAGATGTGCAGATATGGATGTCT 662
Qy 391 ValLeuGlnAenSerLysGluGluAlaLeuAenAenAlaPheLysGlyAspAenThr 410
Db 663 GTACTGCGAAGACAGGAGAGAGAGCTCTGAACAAACGCGCTTTAAGGGTGTGACAACT 722
Qy 411 GlyGluAenAsnIleVal 416
Db 723 GGAGAAATAAATCATCGTC 740
```

RESULT 13  
CN537253 738 bp mRNA linear EST 29-APR-2004  
LOCUS CN537253

## DEFINITION

UI-M-HS0-cqk-j-11-0-UI.r1 NIH\_BMAP\_HS0 Mus musculus cDNA clone

IMAGE:30672898 5', mRNA sequence.

CN537253

VERSION CN537253.1 GI:46865409

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 738)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

## FEATURES

Location/Qualifiers

1..738

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:30672898"

/tissue\_type="Upper Head"

/dev\_stage="embryo 9.5 - 10.5 dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH\_BMAP\_HS0"

Note="Organ: Upper Head; Vector: pYX-Asc; Site\_1: EcoR I;

Site\_2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is CGAACTGAAT. This library was created for the University

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH)."

## ORIGIN

## Alignment Scores:

Pred. No.: 6.05e-79 Length: 738  
Score: 1237.00 Matches: 235  
Percent Similarity: 96.75% Conservatives: 3  
Best Local Similarity: 95.53% Mismatches: 7  
Query Match: 23.46% Indels: 1  
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN537253 (1-738)

```
Qy 329 SerValLysAenGlyPheLeuThrIleSerHisGlyThrAlaAenArgProProAlaLys 348
Db 3 TCCGTTAAATATGGCTCTCTCACCATCTCCACCGCACTGCCACCGCCACCTGCCAAG 62
Qy 349 LeuAenLeuLeuThrCysGlnValLysThrAenProGluGluLysLysCysPheAspLeu 368
Db 63 CTCACCTGCTTAACCTGCCAGGTGAAGACCAATCTCAGGAGAAAGAGTGTTCGACCTC 122
Qy 369 IleSerHisAspArgThrTyrHisPheGlnAlaGluAenGluGlnGluCysGlnIleTrop 388
Db 123 ATATCATGATGACGAGCGTACCACTTCACAGCGGAAGACGAACGAAGATGTCAGATATGG 182
```

```
Qy 389 MetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAep 408
Db 183 ATGCTGTACTGCAGAAACAGCAAGAAAGAGCTCTGAACAACGCCCTTTAAGGGTGTATGAC 242
Qy 409 AsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleSerGluValGln 428
Db 243 AACACTGGAGAAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGGAGGTGCAG 302
Qy 429 ArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTirpleuSer 448
Db 303 AGGATGACGGGCAATAGACGTGTGCTGCCACTGTGGGGCACCAGATCCGACGTGGCTCTCT 362
Qy 449 ThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyVal 468
Db 363 ACCAACCTGGGCATCTGACTTGATCGAGTGTCTGGGATCCACCGGAGCTGGGGTT 422
Qy 469 HisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeu 488
Db 423 CATTACTCCAGATGACGTCCCTGCATTAGATGATTGGGAACGTCTGAGCTTGTGCTT 482
Qy 489 AlalysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGlu 508
Db 483 GCCAAGAAATATCGGGAATGCGAGGCTTTAATGAAATTATGGAGTGTGCTTACCGTCTGAG 542
Qy 509 AspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAla 528
Db 543 GACCCAGTCAAGCCCAACCCAGGAGTGCATGATGTCGAGGAGGATCATCATCAGCC 602
Qy 529 LysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHis 548
Db 603 AAGTACATGAGAGAGAGATATGACCGGAAAAGCATGCCGACCCGCGGCGAAGCTCCAC 662
Qy 549 SerLeuCysGluAlaValLysThrArgAspIlePhe-GlyLeuLeuGlnAlaTyrAlaAs 568
Db 663 AGCCTTTCCGAGGGCGCGTCAAGACACAGACATTTTTTNGGGTTACTCCAAAGCTTATGCTGA 722
Qy 568 pGlyValAspLeuThr 573
Db 723 TGGTGTGACCTGACA 738

CF539222 737 bp mRNA linear EST 12-SEP-2003
IMAGE:30535336 5', mRNA sequence.
CF539222
CF539222.1 GI:34591591
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousef1.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1.737
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
```

```
/clones="IMAGE:30535336"
/tissue_types="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GH0"
/note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAACTGAAT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
```

ORIGIN

Alignment Scores:			
Pred. No.:	8.42e-79	Length:	737
Score:	1235.00	Matches:	238
Percent Similarity:	97.17%	Conservative:	2
Best Local Similarity:	96.36%	Mismatches:	5
Query Match:	23.42%	Indels:	2
DB:	7	Gaps:	0

US-09-914-042-1 (1-1006) x CF539222 (1-737)

Qy	330	VallysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu	349
Db	1	GTTAAATAATGGCTTCCTCACCATCTCCACGGCACTGCCAACCGGCCACCTGCCAAGCTC	60
Qy	350	AsnLeuLeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIle	369
Db	61	AACCTGCTTAACCTGCCAGGTGAGACCCATCTGAGGAGAGAGAGTGTTCGACTCATATA	120
Qy	370	SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrrMet	389
Db	121	TCACATGACAGGAGCTTACCACCTTCCAAGCGGAGACGAACAAGAATGTTCAGATATGGATG	180
Qy	390	SerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPheLysGlyAspAsn	409
Db	181	TCTGTACTGCAGAACACGAAAGAGAGCTCTGAACAACGCCCTTTAAGGGGTGATGACAAAC	240
Qy	410	ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg	429
Db	241	ACTGAGAGAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGGAGGTGCAGAGG	300
Qy	430	MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTirpleuSerThr	449
Db	301	ATGACGGGCAATACGTGTGCTGCGACTGTGGGCGCACAGATCCGACGTGCTCTCTACC	360
Qy	450	AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis	469
Db	361	AACCTGGGCATCTTCGACTTGATCGATGCTCGGATGCCACCGGAGCTGGGGTTCAT	420
Qy	470	TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla	489
Db	421	TACTTCCAGGATGCACTCCCTGACATTAGATGATTATGGGAGCTCTGAGCTTCTGCTGCC	480
Qy	490	LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp	509
Db	481	AAGAATATCGGGAATCGAGGCTTTAATGAAATTATGGAGTGTTCCTACCTGCTGAGGAC	540
Qy	510	SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys	529
Db	541	CCAGTCAAGGCCAACCCAGGCGAGTGCATGATTTCAGAGGAGGACTACATCACAGCCCAAG	600
Qy	530	TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer	549

US-09-914-042-1 (1-1006) x BU322224 (1-861)

QY	200	GlutyrLeuLeuLeuLysValAsnGluIleLysIleLysGlyValAspLeuLeuGlnAen	219
Db	104	CAGTACCTGCTGGAAGTCAACGAATCAAGATTAAAGAGAGTGCCTTTCAGAAC	163
QY	220	LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu	239
Db	164	CTGATCAGTACTTTCATGCTCAATGCAATTTCTTCAGGATCGATTAAAGCAGTTGAA	223
QY	240	SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln	259
Db	224	AGTCTCAAGCCTTCAATGGAGACACTCTCAACAGACCTTTATACGATAAAACAAGCACAG	283
QY	260	AspGluGluAArgGlnLeuIleGlnLeuAArgAspIleLeuLysSerAlaLeuGlnVal	279
Db	284	GATGAAGAAAGAACAATTAACGCAGCTCAGAGATATTTTAAATCTCGCACTCCAGGTT	343
QY	280	GluGlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGln	299
Db	344	GATCAGAAGAGGANTTCTCAGATTGTCAGAGTACAGCTTACAGTTTACATCAGCCTCAG	403
QY	300	GlyAenLysGluHisGlyThrGluAArgAsnGlySerLeuTyrLysLysSerAspGlyIle	319
Db	404	CGAAACAAGGAACATGGCACTGAACGAATGGAGTCTGTACAAGAAAGTGAATGGAATC	463
QY	320	ArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHis	339
Db	464	AGAAAGTGTGCGAGAAAAGGAGTGTCTAGTTTAAAAACGGTTTTCTTACAATTTCCCAT	523
QY	340	GlyThrAlaAenArgProAlaLysLeuAenLeuLeuThrCysGlnValLysThrAen	359
Db	524	GGTACAGCTAACCGACCTCCAGCAAGCTCAATCTGTTAACTGCCAAGTGAACAACAAAC	583
QY	360	ProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAla	379
Db	584	CCAGAGGAGAAAAAATGTTTGACCTCATATCATGACAGGACATACCACATTCCAAGCA	643
QY	380	GluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAenSerLysGluGluAla	399
Db	644	GAAGATGAACAGGAAATGTCAAAATATGGACATCTGTTCTCAAAACAGCAAGGAGGANGCT	703
QY	400	LeuAenAenAlaPheLysGlyAspAspAsnThrGlyGluAenAenIleValGlnGluLeu	419
Db	704	TTAAATATGATTCACAGGAGATGATAACACAGGAGAAATATATATGTTCCAGGAACTG	763
QY	420	ThrLysGluIleLleSerGluValGlnArgMetThrGlyAenAspValCysCysAspCys	439
Db	764	ACAAAAGAGATTATCTGAAGTCCAAAGCATGACTGGAAATGATGCTGGTGTGCTGCTGT	823
QY	440	-GlyAlaProAspProThrTrpLeuSerThrAen	450
Db	824	GGGAGACAAGATTCTACCTGGGTGCCACCAAT	857

Search completed: August 4, 2005, 13:57:03  
Job time : 7518 secs

240	SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln	250
241	...:::...	
242	...	
243	...	
244	AGTCTCAAGCCTTCAATGGAGACACTCTCAACAGACTTTATACGATAAAACAAGCACAG	283
245	...	
246	...	
247	...	
248	...	
249	...	
250	AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal	279
251	...	
252	...	
253	...	
254	GATGAAGAAAGACAAATAACGCAGCTCAGAGATATTTTAAAACTCGCACTCCAGGTT	343
255	...	
256	...	
257	...	
258	GluGlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGln	299
259	...	
260	...	
261	...	
262	...	
263	GATCAGAAGAAGGATTTCTCAGATTCGTACAGGTACAGCTTACAGTTTACATCAGCCTCAG	403
264	...	
265	...	
266	...	
267	...	
268	...	
269	...	
270	GlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIle	319
271	...	
272	...	
273	...	
274	...	
275	...	
276	...	
277	...	
278	...	
279	...	
280	...	
281	...	
282	...	
283	...	
284	...	
285	...	
286	...	
287	...	
288	...	
289	...	
290	...	
291	...	
292	...	
293	...	
294	...	
295	...	
296	...	
297	...	
298	...	
299	...	
300	GlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIle	319
301	...	
302	...	
303	...	
304	...	
305	...	
306	...	
307	...	
308	...	
309	...	
310	...	
311	...	
312	...	
313	...	
314	...	
315	...	
316	...	
317	...	
318	...	
319	...	
320	ArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHis	339
321	...	
322	...	
323	...	
324	...	
325	...	
326	...	
327	...	
328	...	
329	...	
330	...	
331	...	
332	...	
333	...	
334	...	
335	...	
336	...	
337	...	
338	...	
339	...	
340	...	
341	...	
342	...	
343	...	
344	...	
345	...	
346	...	
347	...	
348	...	
349	...	
350	...	
351	...	
352	...	
353	...	
354	...	
355	...	
356	...	
357	...	
358	...	
359	...	
360	...	
361	...	
362	...	
363	...	
364	...	
365	...	
366	...	
367	...	
368	...	
369	...	
370	...	
371	...	
372	...	
373	...	
374	...	
375	...	
376	...	
377	...	
378	...	
379	...	
380	...	
381	...	
382	...	
383	...	
384	...	
385	...	
386	...	
387	...	
388	...	
389	...	
390	...	
391	...	
392	...	
393	...	
394	...	
395	...	
396	...	
397	...	
398	...	
399	...	
400	...	
401	...	
402	...	
403		

524	GGTACAGCTAACCGACCTCCAGCAAGCTCAATCTGTTAACCTGCCAAGTGAACAACAAC	583
QY	ProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTrpHisPheGlnAla	379
DB	CGAGGAGGAGAAAAAATGTTTGGACCTCATATCATCATGACGAGGACATACCACATTTCCAAGCA	643
OV	GluAspGluGlnGluCysGlnIleTromMetSerValLeuGlnAsnSerLysGluGluAla	199

```

07 556 GlnAspGluIleGluCysAlaIleIleIleIleIleIleIleIleIleIleValGlnGluLeu 419
Db 644 GAAGATGAACAGGAATGTCAAAATATTGGACATCTGTTCTACAAACACGCAAGGAGGAAGCT 703
QY 400 LeuAenAenAlaPheLysGlyAspAsnThrGlyGluAenAenIleValGlnGluLeu 419
Db 704 TTAAATAATGCATTCACAGGAGATGATAACACAGGAGAGAAAATAATATGTC CAGGAAC TG 763
QY 420 ThrLysGluIleIleSerGluValGlnArgMetThrGlyAenAspValCysCysAspCys 439
Db 764 ACAAAAGAGATTATATCTGAAGTCCAAAGCATGACTCGAAATGATGTGTGTGTA CTGT 823
QY 440 -GlyAlaProaspProThrTrpLeuSerThrAan 450
Db 824 GGAGACACAAGATTCTACCTGGGTGCCACCAAAT 857

Search completed: August 4, 2005, 13:57:03
Job time : 7518 secs

```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:34:52 ; Search time 8647 Seconds  
(without alignments)  
4749.391 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPERKAGPVPVVFHFIAD 1006

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 45554873 seqs, 20411521753 residues

Total number of hits satisfying chosen parameters: 91109746

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US09914042/runat\_04082005\_090634\_2150/app\_query\_faeta\_1.1159  
-DB=Pending\_Patents\_NA\_Main -OPMT=fastap -SUFFIX=rnppm -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042@cgn2\_1 1 5097 @runat\_04082005\_090634\_2150 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents NA\_Main:  
1: /cgn2\_6/ptodata/1/pna/PCTUS1\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/PCTUS2\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US08\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*

29: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq.\*  
44: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US099G\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US103A\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US103B\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US104A\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US104B\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US105A\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US105B\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US106A\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US107A\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US107B\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US107C\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US107D\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US108A\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US108B\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US109A\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US109B\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US109C\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US110\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US6023A\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US6023B\_COMB.seq.\*  
94: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
95: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
96: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
97: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
98: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
99: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
100: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
101: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*

102: /cgn2\_6/ptodata/1/pna/US6032 COMB.seq.\*  
 103: /cgn2\_6/ptodata/1/pna/US6033 COMB.seq.\*  
 104: /cgn2\_6/ptodata/1/pna/US6034 COMB.seq.\*  
 105: /cgn2\_6/ptodata/1/pna/US6035 COMB.seq.\*  
 106: /cgn2\_6/ptodata/1/pna/US6036 COMB.seq.\*  
 107: /cgn2\_6/ptodata/1/pna/US6037 COMB.seq.\*  
 108: /cgn2\_6/ptodata/1/pna/US6038 COMB.seq.\*  
 109: /cgn2\_6/ptodata/1/pna/US6039 COMB.seq.\*  
 110: /cgn2\_6/ptodata/1/pna/US6040 COMB.seq.\*  
 111: /cgn2\_6/ptodata/1/pna/US6041 COMB.seq.\*  
 112: /cgn2\_6/ptodata/1/pna/US6042 COMB.seq.\*  
 113: /cgn2\_6/ptodata/1/pna/US6043 COMB.seq.\*  
 114: /cgn2\_6/ptodata/1/pna/US6044 COMB.seq.\*  
 115: /cgn2\_6/ptodata/1/pna/US6045 COMB.seq.\*  
 116: /cgn2\_6/ptodata/1/pna/US6046 COMB.seq.\*  
 117: /cgn2\_6/ptodata/1/pna/US6047 COMB.seq.\*  
 118: /cgn2\_6/ptodata/1/pna/US6048 COMB.seq.\*  
 119: /cgn2\_6/ptodata/1/pna/US6049 COMB.seq.\*  
 120: /cgn2\_6/ptodata/1/pna/US6050 COMB.seq.\*  
 121: /cgn2\_6/ptodata/1/pna/US6051 COMB.seq.\*  
 122: /cgn2\_6/ptodata/1/pna/US6052 COMB.seq.\*  
 123: /cgn2\_6/ptodata/1/pna/US6053 COMB.seq.\*  
 124: /cgn2\_6/ptodata/1/pna/US6054 COMB.seq.\*  
 125: /cgn2\_6/ptodata/1/pna/US6055 COMB.seq.\*  
 126: /cgn2\_6/ptodata/1/pna/US6056 COMB.seq.\*  
 127: /cgn2\_6/ptodata/1/pna/US6057 COMB.seq.\*  
 128: /cgn2\_6/ptodata/1/pna/US6058 COMB.seq.\*  
 129: /cgn2\_6/ptodata/1/pna/US6059 COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5253	99.6	5711	41	US-09-949-002-75
2	5253	99.6	5711	41	US-09-949-003C-427
3	5253	99.6	5711	66	US-10-956-157-1144
4	5253	99.6	5711	120	US-60-507-511-1144
5	5253	99.6	5756	34	US-09-770-174-4023
6	5240	99.4	5720	85	US-60-172-360-28138
7	5236	99.3	5810	90	US-60-213-359-1582
8	5236	99.3	5810	97	US-60-278-232-6642
9	5210	98.8	6389	26	US-09-572-411-6530
10	5152	97.7	5670	49	US-10-170-235-8470
11	5146	97.6	5670	115	US-60-452-680-5751
12	4664	88.5	3025	41	US-09-949-002-139
13	4664	88.5	3025	41	US-09-949-003C-925
14	4568.5	86.6	5427	49	US-10-170-235-8660
15	4562.5	86.5	5427	115	US-60-452-680-5750
16	4152.5	78.8	5619	48	US-10-144-771-17053
17	4152.5	78.8	5619	106	US-60-360-207-17053
18	3029	57.4	3775	1	PCT-US02-36151-103
19	3026.5	57.4	6110	62	US-10-796-280-303
20	3026.5	57.4	6110	67	US-10-995-561-189
21	3026.5	57.4	6110	126	US-60-568-219-103
22	3026.5	57.4	6212	62	US-10-796-280-302
23	3026.5	57.4	6212	67	US-10-995-561-188
24	3026.5	57.4	6212	126	US-60-568-219-102
25	3024.5	57.4	3946	1	PCT-US02-36151-96
26	3006	57.0	3974	50	US-10-210-281-73
27	2786	52.8	4304	48	US-10-144-771-9150
28	2786	52.8	4304	106	US-60-360-207-9150
29	2556.5	48.5	4997	2	PCT-US03-30720-325
30	2555.5	48.5	5033	22	US-09-488-725A-1427
31	2555.5	48.5	5033	29	US-09-653-450A-14
32	2555.5	48.5	5033	51	US-10-258-898A-1427
33	2555.5	48.5	5033	51	US-10-286-897-1427
34	2380.5	45.1	5445	102	US-60-324-185-16737
35	2371.5	45.0	5582	28	US-09-625-102-1606

36 2371.5 45.0 5582 47 US-10-062-674-1606  
 37 2166.5 41.1 4719 115 US-60-453-050-2647  
 38 2166.5 41.1 4719 115 US-60-453-135-2647  
 39 2166.5 41.1 4719 116 US-60-466-412-2647  
 40 2164.5 41.0 4719 49 US-10-170-235-24226  
 41 2089.5 39.6 3812 48 US-10-104-047-981  
 42 2089.5 39.6 4050 55 US-10-490-605A-1  
 43 2089.5 39.6 4050 55 US-10-490-605A-1  
 44 2089.5 39.6 4050 62 US-10-777-860A-38  
 45 2044.5 38.8 4256 49 US-10-170-235-28087

## ALIGNMENTS

## RESULT 1

US-09-949-002-75  
 ; Sequence 75, Application US/0949002  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; FILE REFERENCE: CL000790  
 ; CURRENT APPLICATION NUMBER: US/09/949,002  
 ; CURRENT FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: 60/231,401  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 10823  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 75  
 ; LENGTH: 5711  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-002-75

## Alignment Scores:

Pred. No.: 3.37e-217 Length: 5711  
 Score: 5253.00 Matches: 1003  
 Percent Similarity: 99.70% Conservative: 0  
 Best Local Similarity: 99.70% Mismatches: 3  
 Query Match: 99.62% Indels: 0  
 DB: 41 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-949-002-75 (1-5711)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
 Db 341 ATGCCGACACAGATCTCCGTGCGAATTCGTGCCGAGACCCATGAGGACTCAAGCGC 400

Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40  
 Db 401 CCCAGCGCCTCCAGCTTCACCCCGCAGCGAGTCCGGAACACTGTGGCGGCATC 460

Qy 41 GluAlaAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
 Db 461 GAGGAGGCTTGGACGTGGACCGGATGGTCTTTACAAATGAAGAAATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLysAlaHisValGluAsnGluGluThrGlnAlaLeuGlu 80  
 Db 521 ATCAACAGCTCTGGCTGGCTGCGTACGTGGAATAATGAAGAGCAGTACACCGGCTCTGGAG 580

Qy 81 LysPheGlyGlyAenCysValCysArgAspProPheLeuGlySerAlaPheLeuLys 100  
 Db 581 AAGTTTGGCGCACTCTGTATGCAGAGATGACCCAGATTTAGGAAGTGCCTTCTTGAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 Db 641 TTCTCAGTGTTCAAAGGAGTTGACAGCACTTTTCAAAACCTGATTCAAGATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
 Db 701 AACATAATCTCTTCCCTTTGGACAGTTTGTCTGAAGGGGACCTTGAAGAGGAGTGAAGGG 760

QY 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
DB 761 GATCTGAAAAAGCCCTTTTGATAAAGCTTGGAGGACTATGAAACAAAATAACCAAGATA 820  
QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180  
DB 821 GAAAGGAGAAAGGAAACACCCCAAGCTCCATGGGATGATTCCGAGCTGAATAAGCCGA 880  
QY 181 AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
DB 881 GCGGAAATTTGCCGAAGAGATGGAAGAGAGAGCGCTTCTCCAGCTACAGATGTGCCAG 940  
QY 201 TyrLeuLeuLysValAsnGluLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220  
DB 941 TATCTGCTCAAGGTCAAGCAATCAAGATTAAGGAGGAGTAGATTTTACTTTCAGAACTG 1000  
QY 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240  
DB 1001 ATCAATACTTTCATGCCCAATGCAATTTTTTTCAGGATGGAATCAAGCCGTGGAAGC 1060  
QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
DB 1061 CTCAAACCTTCATTGAACGCTGTCTACGGATCTTCACAGATCAACAGGCCCCAGGAT 1120  
QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
DB 1121 GAAGAAAGAGGAGCAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCGAGGTTGAA 1180  
QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
DB 1181 CAGAAAGAGGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTATACATCAGGCTCAGGGA 1240  
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
DB 1241 AACAGGAACATGGACCGGAGCGGAAACCGGAGCTCTACAGAGAGTGACGGGATCCGA 1300  
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
DB 1301 AAAGTGTGCGCAAAAGGAAATGTTTCAGTTAAATAATGGTTTTCTGACCATATCCCATGGT 1360  
QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
DB 1361 ACCGCTAACCGGCGCTCTGCAAGCTCAACCTGCTTAACCTGCGAGTGAAGACCAACCTT 1420  
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
DB 1421 GAGGAGAGAAGTGTCTTGACCTTATTTACATGACAGACTTACCATCTTCAAGCTGAA 1480  
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu 400  
DB 1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAAAGAAAGCTTTA 1540  
QY 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
DB 1541 AACAAATGCAATTAAGGGGATGACAATACTGGAGAAATAAATCATCGTCCAAAGAACTGACA 1600  
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
DB 1601 AAGGAGATCACTCAGAGTGCAGAGGATGACGGCAATGACCTCTGTGTGACTGTGGG 1660  
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
DB 1661 GCGCAGATCTTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTC 1720  
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
DB 1721 GGAATCCACCGAGAGCTGGGGTTCATTATTCAGGATGCAGTCCCTGACCTTAGATGTA 1780  
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
DB 1781 CTGGGAACATCTGAGCTGCTGCCCAAGAAATTTGGAAATGCAAGGCTTTAATGAGATC 1840  
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520

DB 1841 ATGGAATGTTCCCTACCAGCTGGAGACTCAGTCAAAACCAACCCAGGCGAGCATGAT 1900  
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
DB 1901 GCAAGAAAGGACTACATCAGACCCAGTACATCAGAGGAGATACGCAAGGAGNAGAC 1960  
QY 541 AlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
DB 1961 GCGGATAACGCGCGGAAGCTTCACAGTCTTTGCGAGGCGGTCAAAACAGAGATATTTTT 2020  
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
DB 2021 GGATTTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAATAAATCCCACTGGCCAA 2080  
QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
DB 2081 GGACATGACCGGATGAAACGCGCTCCACCTTGCACTGAGTCCGCTGATCGAACCTCT 2140  
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
DB 2141 CTTCACTTTGTAGACTTTTTTAGTTTCAGAAACAGTGGGAACCTGGATAAACAGACAGGAAA 2200  
QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
DB 2201 GGCAGCAGACCCCTGCACTACTGCTGCTGACCCGACAAATGCCGAGTGGCTCAAGTTGCTC 2260  
QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660  
DB 2261 CTGGGGGGAAGGCTCCATCGAGATGCAAAACGAGTCCAGGAGACTCCGCTGGACATT 2320  
QY 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
DB 2321 GCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGTGACCCAAAGCTTATCTGGAAGA 2380  
QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
DB 2381 TTTAATTTCTACGTTACGTTGAATATGAAATGGCGACTACTCCAGAAAGACCTTGATGAA 2440  
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
DB 2441 AGTGAATGACACATGATGATGAGAAATTCGAGCCAGTCCCAACCGCGGGAGACCGGCC 2500  
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
DB 2501 ATCAGCTTCTACCAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560  
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
DB 2561 GATGCTGCAAACTTGGCAGGAGAGCAGAGGCGCTTTCATGCCCAGCATCTTGCAAGAT 2620  
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
DB 2621 GAGACTTACGAGAGCCCTCTGAGTGGCAGCCACCTCCCGCCAGCTTCGAGCCCCAGC 2680  
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
DB 2681 ACCACCGCGCCCGCCGCTTCTCCACGGAATGTTGCAAAAGTTTCAGACAGCCTCTCT 2740  
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyLysArgGlnArgSer 820  
DB 2741 GCTAACACCTTGTGGAAGACAAACTCTGTAAGTGTGGACGGTGGGAAGCGCGGAGCATCT 2800  
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840  
DB 2801 TCGTCAGATCCGCGAGCTGTCTCCACCGCTGCCCTCTTTCGGGTGACATCTACCANT 2860  
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
DB 2861 CCCCTGACCCCGCGCGCCCGCCCGCTTCCCAAGACGCCCGAGCGTAATGGAGCCCTTG 2920  
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880

```
Db 2921 AGCCAGCCGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCACCTCTGCCC 2980
Qy ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCCGCCAGCGCTCCCGCAGAGAAGCCCTGCCCGGGGGCTGCACAAGTCCACC 3040
Qy ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAGGCCCAACGAGAGAGACTGTGGATCTCTCTGCAACGGAAGCTCTG 3100
Qy GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGCTGCTGAGGCCCTCGCAGCCCTCGCACCCTAGGAAGTCGAG 3160
Qy AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAACCC 3220
Qy AspGluLeuThrPheSerGluGlyAspValLlelleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACTTCTCCAGGGGATGTGATCATCTGTGGACGGGAGGAGACCGAG 3280
Qy TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATTGCCACATTGATGGAGATCCTGTGTGCAAGGCGCATTCCTCGGTGTCAATT 3340
Qy ValHisPheIleAlaAsp 1006
Db 3341 GTGCACCTTTATCGCTGAC 3358

RESULT 2
US-09-949-003C-427
; Sequence 427, Application US/09949003C
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPYSCHIAITRIC
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000791
; CURRENT APPLICATION NUMBER: US/09/949,003C
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,446
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 74065
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 427
; LENGTH: 5711
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-949-003C-427

Alignment Scores:
Pred. No.: 3,37e-217 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 41 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-949-003C-427 (1-5711)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCCGACCAAGATCTCGTGTGGAAATCTGTGGCCGAGACCCATGAGACTACNAGGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCAGCGCTCCAGCTTCACCACCCGCGAGCGGAGTCCGGAACACTGTGGCGGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACGTGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
```

```
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGGAATAAGAGCAGTACACCCAGGCTCTGGAG 580
Qy LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCACAGATTTAGGAAGTGCCTTCTCTGAAG 640
Qy PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTTCAAAAGGAGTTGACAGCACTTTTCAAAAACCTGATTTCAGAAATATGAAC 700
Qy AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATATATCTCTTCCCTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTGAAGGG 760
Qy AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 761 GATCTGAAAAAGCCCTTTTGTATAAGCTTGAAGGACTATGAACAACAAAATAACCAAGATA 820
Qy GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTTCGAGACTGAAATTAAGCGGA 880
Qy AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAATTTGCCGAAGAGATGGAAGAGAGAGGGCGCTTCTTCAGACTACAGATGTGCGAG 940
Qy TyrLeuLysValAsnGluIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGCTGAAGTCAACGAAATCAAGATTAAGGAGTAGATTTACTTTCAGAACTCTG 1000
Qy IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATACTTTCATGCCCAATGCAATTTTTCAGATGGACTCAAGGCGCTGGAAGC 1060
Qy LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTTCCATTGAACCGCTGTCTACGGATCTTCACAGCATCAACAGGCCCCAGGAT 1120
Qy GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCGAGTTGAA 1180
Qy GlnLysGluAspSerGlnIleAspGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGGA 1240
Qy AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAAAGGACATGGGACCGAGCGGAACGGCAGCCTCTACAAGAGAGAGTACGGGATCCGA 1300
Qy LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAAGAAATGTTTCAAGTAAAAAATGTTTTCTGACCATATCCCATGGT 1360
Qy ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAGAGCTCAACCTGTCTAACCTGCCAGGTGAAGCAACCCCT 1420
Qy GluLysLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAAGAGTCTTTGACCTTATTTTACATCAGACAGACTTACCACCTTTCAAGTGAA 1480
Qy AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATGAACAGGAATGTCAAAATATGATGTCTGTGTCGCAAAAATAGCAAAAGAAAGAGCTTTA 1540
Qy AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACATGCTATTAAGGGGAGTACAAATCTGGAGAAATATACATCTGTCACAGAACTGACA 1600
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
```

Db 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGACGTGTGGG 1660  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1661 GCGCAGATCTACATGCTTTCCACCAACCTGGGCATCTCTGACCTGCATCGAGTGTTC 1720  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrIleuAspVal 480  
Db 1721 GGAATCCACCGAGAGCTGGGGTTCATTAATTCAGGATGCAGTCCCTGACCTTAGATGA 1780  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 CTGGACATCTGAGCTCTGCTGCCCAAGATATTGGGAATGACAGGCTTTAATGAGATC 1840  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1841 ATGGAATGTTGCTTACCAGCTGAGACTCAGTCAAAACCCCAACCCAGGCGAGCATGAAT 1900  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHis 540  
Db 1901 GCAAGAAAGGACTACATCACAGCCAAGTACATCGAGAGGAGATACGCAAGGAAGACAC 1960  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1961 GCGATACCGCGGAGCTTCACAGTCTTTGCGAGGCCGTCAAAACCGAGAGATATTTT 2020  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2021 GGATTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAA 2080  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2081 GGACATGAGCGGAGTAAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2141 CTTACATGTGAGACTTTTATGTTAGTACAGACAGTGGAAACCTGGATAAACACAGAGGAAA 2200  
Qy 621 GlySerThrAlaLeuHisTyrCysLeuSerThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2201 GGCAGCAGCCCTGCATCTACTGCTGCTGACCGACCAATGCGAGTGCTCAAGTGTCTC 2260  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2261 CTGCGGGGGAAGGCTCTCCATCGAGATAGCAACAGAGTCAGGAGACTCCCTCGACATT 2320  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2321 GCCAAGCGCCCTCAAGACGACGACTGTGAGAGAGCTGTGACCAAGCCCTTATCTGGAAGA 2380  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2381 TTTAATTTCTACGTTTCAGTTGATATGAATGGCGACTACTCCACGAGAGACTGTGATGAA 2440  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2441 AGTGATCAGCATGAGTGAATAATGAGCCAGTCCCAACCGCGGGAAGACCGGCC 2500  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2501 ATCAGCTCTTACCAGCTGGGCTCAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2561 GATGCTGCNAACCTTGCACAGAGAGACGAGGAGGCTTTCATGCCAGCATCTTGCAGAT 2620  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2621 GAGACTTACGAGAGCCCTCTCAGTGGCAGCCACCTCCCGCCAGCTGCGAGCCCGCAGC 2680  
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2681 ACCACGAGCGCCCCCGCTTCTCCACGGAAATGTTGGCAAAAGTTTCAGACAGCCTCTCT 2740

Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyLysArgGlnArgSer 820  
Db 2741 GCTAACACCTGTGGAGACAAACTCTGTAGTGTGGACGCTGGGAAGCCGCGAGCATCT 2800  
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840  
Db 2801 TCGTCAGATCCGCGAGCTGTCCATCCACCGTGCCTCTTTCGCGTGACATCTACCAAT 2860  
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
Db 2861 CCCCTGACCCCAAGCGCCGCCACCTGTTGCCAAGACGCCCGAGGCTTAATGGAGCCTTG 2920  
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880  
Db 2921 AGCCAGCGAGCAAGCTGCTCCCGCTGGGATCTCACAGATCAGGCCCCCACTCTGCC 2980  
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900  
Db 2981 CCACAGCGCCCAAGCGCCCTCCCGCAGAAGAGCTGCGCGGGGGCTGACAAAGTCCACC 3040  
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920  
Db 3041 CCATGACCAACAAGGCCAACCGAGAGGACTGTGGATCTCTCTGCAACGGAAGCTCTG 3100  
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940  
Db 3101 GGTCTCTGTCCAATGCTATGCTCTGTCAGCCCTTCACCATGCTAGGAAGTCGACAG 3160  
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960  
Db 3161 GCAACCAAGTTGAAGCCCTTAAGCGGGTGAAGCGCTCTATAACTGTGTGGCTGACAACCC 3220  
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980  
Db 3221 GATGAGCTCAGCTTCTCGAGGGGATGTGATCATCTGTGACGGGAGAGACCGAGGAG 3280  
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000  
Db 3281 TGGTGGATGGCCACATTTGATGGAGATCTCTGTCGAAAGCGCATTCCTCCGCTGTCTTT 3340  
Qy 1001 ValHisPheIleAlaAsp 1006  
Db 3341 GTGCACCTTTATCGCTGAC 3358  
RESULT 3  
US-10-956-157-1144  
; Sequence 1144, Application US/10956157  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956.157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1144  
; LENGTH: 5711  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-1144  
Alignment Scores:  
Pred. No.: 3,37e-217 Length: 5711  
Score: 5253.00 Matches: 1003  
Percent Similarity: 99.70% Conservative: 0  
Best Local Similarity: 99.70% Mismatches: 3  
Query Match: 99.62% Indels: 0  
DB: 66 Gaps: 0  
US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)

Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
Db 341 ATGCCGGACCAAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTCAAGAGCG 400  
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
Db 401 CCCAGGCTCCAGCTTCAACCCCGCACGGCGCAGTGC CGGACACTGTGGCGGCCATC 460  
Qy 41 GluGluAlaLeuAppValAspArgMetValLeuTyrIysMetLysLysSerValIysAla 60  
Db 461 GAGGAGGCTTTGGACCTGGACCGGATGTTCTTTACAAAATCAAGAAATCCGTGAAGCA 520  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGCAGTACACCCAGGCTCTGGAG 580  
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
Db 581 AAGTTTGGCGGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTCCGTCTCTGAAG 640  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db 641 TTCTCAGTGTTTACAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAAGATATGAAC 700  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140  
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTGAAGGGGACCTGAAAGGAGTGAAGGG 760  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 761 GATCTGAANAAGCCTTTTGATTAAGCTTGGAGGACTATGAACAAAATTAACCAAGATA 820  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db 821 GAAAGGAGAAAAAGAAACACGCCAAGCTCCATGGGATGATTGGACTGAAATAAGCGGA 880  
Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
Db 881 GCGGAAATTTGCGAAGAGATGGAAAGAGAGGCGCTCTTCCAGCTACAGATGTGCGAG 940  
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 941 TATCTGTGAAGTCAACGAAATCAAGATTAAAGAGGAGTAGATTACTTCAGNATCTG 1000  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 1001 ATCAAAATACTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAAAGCCGTGAAAGC 1060  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1061 CTCAAACCTTCATTGAACCGCTGTCTACGATCTTCACAGATCAAAACAGGCCCCAGGAT 1120  
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1121 GAAGNAAGAGCGCAGTTGATACAGCTTCAGNATATTTTGAATTCGCAATTGCAGGTGAA 1180  
Qy 281 GlnLysGluAppSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 1181 CAGAAAGAGGACTCCCAAATTCGTCAAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 1241 AACAAAGNAACATGGGACCGAGCGGAACCGGACGCTCTCAAGAAAGAGTGACGGGATCCGA 1300  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1301 AAAGTGTGGCAGAAAGGAAATGTTCAGTTAATAATGTTTTTCTGACCCATATCCATCGT 1360  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGCTTAACCTGCCAGGTGAAGACCAACCT 1420

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1421 GAGGAGAAAGAGTGTCTTTTGACCTTATTTTACATGACAGAACTTACCACTTTCAAGCTGAA 1480  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1481 GATCAACAGGAATGTCATAATATGATGCTGTGCTGCACAAATAGCAAAAGAAAGCTTTA 1540  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGlnLeuThr 420  
Db 1541 AACAAATCATTTAAGGGGATGACAATACTGGAGAAAATAACATCGTCCCAAGAACTGACA 1600  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1601 AAGAGATCATCTCAAGATGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGGG 1660  
Qy 441 AlaProaspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1661 GCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTCTGACCTGCTCGAGTGTTC 1720  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1721 GGAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGGATGCAGTCCCTGACCTTAGATGTA 1780  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 CTGGGAACATCTGAGCTGTCTGCCAGNATATTGGGAATGCAGGCTTTAATGAGATC 1840  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1841 ATGGAATGTTGCCCTACAGCTGAGGACTCAGTCAAAACCCACCCAGGCGAGCATGAAT 1900  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCCAGAGGAGATACGCAAGAAAGAAC 1960  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1961 GCGGATTAACGGCGGAGAGCTTCACAGCTTTTTCGAGGCCCTCAAAACGAGAGATATTTT 2020  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2021 GGATTTGCTCCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAA 2080  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2081 GGACATGAGCGGATGAACCGCCCTCCACCTTCAGTCAGTCCGAGTCCGAGTCCAACTCT 2140  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2141 CTTCACATTTGTAGACTTTTATTAGTTTCAGAACAGTGGGAACCTTGGATTAACACAGAGGAA 2200  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2201 GGCAGCACAGCCCTGCACCTACTGCTGCTGACCCGACCAATGCCGAGTGCCTCAAGTTGCTC 2260  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660  
Db 2261 CTGGGGGGAAGGCTTCATCGATAGCAACAGATCAGAGTCCAGAGAGACTCCCGCTGACAT 2320  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2321 GCCAAGCCCTCAAGCAGCAGCAGCTGTGAGAGGCTGTGTGACCCCAAGCCTTATCTGGAAGA 2380  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2381 TTTAATTTCTACGTTTCAGTTGAATATGAATGGCGACTACTCCACGAAGACCTCGATGAA 2440  
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2441 AGTATGACGACATGGATGGAAATTCAGCCCCAGTCCCAACCCGCGGGAAGACCGGCC 2500  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740

```
Db 2501 ATCAGCTCTTACCAGCTGGCTCCACACAGCTTCAAGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAen 760
Db 2561 GATGCTGCAACCTTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2620
Qy 761 GluThrTyGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAAsnProSer 780
Db 2621 GAGACTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2740
Qy 801 AlaAenThrLeuTrpLysThrAenSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAen 840
Db 2801 TCGTCAGATCGCAGAGCTGTCATCCACCGCTGCCCTCTTGGCGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCAGCCCGCCCGCCCGCTTGCAGAGAGAGAGAGAGAGAGAGAT 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCAGCGCTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACGTACCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCAATGCTATGCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTAAGCGGCTGAAGCGCTCTATACTGTGTGCTGACACCCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValLleValleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3280
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATGGCCACATGATGGAGATCCTGTGTCGAAAGGCGCATCCCCGTGTCAATT 3340
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3341 GTGCACCTTATCGCTGAC 3358
```

## RESULT 4

```
US-60-507-511-1144
; Sequence 1144, Application US/60507511
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: AM 101081
; CURRENT APPLICATION NUMBER: US/60/507,511
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 203623
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1144
```

```
; LENGTH: 5711
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-507-511-1144
Alignment Scores:
Pred. No.: 3,376-217 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservatives: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
Gaps: 120
US-09-914-042-1 (1-1006) x US-60-507-511-1144 (1-5711)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyLysAla 20
Db 341 ATGCGGACCAAGATCTCGTGTCCGAAATTCGTGGCCGAGACCCATGAGGACTTACAGGGG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGGCTCCAGCTTCCACCCGCGCAGCGCGAGTCCCGACACACTGTGGCGGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGCTTGGACGCTGGACCGATGCTTCTTACAAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGGTGGCTCAGGTGAAAATGAGAGACAGTACACCGAGCTCTGGAG 580
Qy 81 LysPheGlyGlyAenCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGACAGATGACCCAGATTTAGGAAGTGCCTTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuIleGlnAenMetAen 120
Db 641 TTCTCAGTGTTTACAAAGGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAATATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTCTGAAGGGGGAGCTGAAAGAGGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyGluThrLysIleThrLysIle 160
Db 761 GATCTGAAAAGCCTTTTGATAAGCTTTGGAAGGACTATGAAACAAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
Db 821 GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTTGCGAAGAGATGGAAGAGAGAGAGGCGCTTCTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLeuLysValAenGluIleLysIleLysLysGlyValAspLeuGlnAenLeu 220
Db 941 TATCTGCTGAAGTCAACGAAATCAAGATTAAGGAGGAGTAGATTTACTTCAGAATCTG 1000
Qy 221 IleLysTyPheHisAlaGlnCysAenPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTCATGTCGCAATGCAATTTTTTTCAGGATGGAATCAAGAGCCGTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAen 260
Db 1061 CTCAAACCTTCCATTTGAACCCCTGTCTACGATCTTTCACAGATCAAGAGCCCGCAGAT 1120
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGAGTGTGATACAGCTTCGATATTTTGAATTCGATTCGAGGTGAA 1180
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTySerLeuHisGlnProGlnGly 300
```



Db 1181 CAGAAAGGAGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCTCAGGGA 1240  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLeuArg 320  
Db 1241 AACAAAGAACATGGACCGGACCGGACCGCTCTCAAGAAAGAGTGCAGGATCCGA 1300  
Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1301 AAAGTGTGGCAGAAAGGAATGTTCACTTAAATATGGTTTTCTGACCATATCCCATGGT 1360  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1361 ACCGTAACCGGCTCTCTGCAAGACTCAACCTGTCTAACCTGCCAGTGAAGACCAACCT 1420  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1421 GAGGAGAAAGAGTGTCTTGGACTTATTTTCAATGACAGAACTTACCACTTTCAAGCTGAA 1480  
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAAAGAAAGCTTTA 1540  
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1541 AACAAATGCAATTAAGGGGATGACAAATACTGGAGAAAATAACATCGTCCAAAGAACTGACA 1600  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1601 AAGGAGATCATCTCAGAAATGTCAGAGGATGACGGCAATGACGTGTGTGACTGTGGG 1660  
Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1661 CGCCAGATCTCATATGCTTTCACCAACCTGGGCATCTTGACCTGCATCGATGTGTCC 1720  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1721 GGAATCCACCGAGAGCTGGGGTTCATTATTCAGAGATGCAGTCCCTGACCTTAGATGTA 1780  
Qy 481 LeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 CTGGGAACATCTGAGCTCTCTGCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 1840  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1841 ATGGAAATGTTCCTACCAAGCTGAGACTTCAGTCAAAACCCCAACCCAGGCGAGCAATGAA 1900  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1901 GCAAGAAAGGACTACATCACAGCCAAGTACATCGAGAGGAGATACGCAAGGAAGAGCAC 1960  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1961 CGCGATACCGGGCGAAGCTTCACAGTCTTTGCGAGGCGCTCAAAACAGAGAGATATTTTT 2020  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2021 GGATTTCTCCAAAGCTTATGCTGATGGTGTGAATCTTACGGAAAAAATCCCACTGGCCAA 2080  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2141 CTTTCACTTGTAGACTTTTATAGTTAGTTCAGAACACAGTGGGAAACCTGGATAAACAGACAGGGA 2200  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2201 GGCAGCACAGCCCTGCACTACTGTCTGCTGACCGCAATGCGAGTGTCTCAAGTTGTCT 2260  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2261 CTGGGGGGAAGGCCCTCCATCGAGATACAAACGAGTCCAGGAGACTCCGCTGGACATT 2320

Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2321 GCCAAGCGCTTCAAGCACGAGACTGTGAGAGTGTGTGACCAAGCCTTATCTGGAAGA 2380  
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2381 TTTTAAATTCACGTTACGTTGAATATGATGGCGACTACTCCACGAAGACCTGGATGAA 2440  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2441 AGTGATCAGACATGATGATAAATTCAGCCAGTCCCAACCGCGGGAAGACCGGCC 2500  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2501 ATCAGCTTCTACCAAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2561 GATGCTGCAACCTTGCNAGGAGAGAGAGGGCTTTTCATGCCAGCATCTTCAGAAAT 2620  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780  
Db 2621 GAGACTTACGAGCGCTCTCAGTGGCAGGCCACTCCCGCCAGCCTGCAGCCCCCAGC 2680  
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2681 ACCACCGAGCGCCCGCTTCTCCACGGAATGTGGCAAGTTCAGACAGCCTCTCTCT 2740  
Qy 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
Db 2741 GCTAACACCTGTGGAAGACAAACTCTGTAAGTGTGGAGCGTGGAGCCGCGCGATCT 2800  
Qy 821 SerSerAspProProAlaValHisProLeuProProLeuArgValThrSerThrAsn 840  
Db 2801 TCGTCAGATCCGCGAGCTGTCCATCCCGCTGCCCTCTCTTCGCGTGACATCTACCAAT 2860  
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
Db 2861 CCCGTGACCCCGCCCGCCCGCTTCCCAAGAGCGCCAGCGTAAATGGAAAGCCTTG 2920  
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880  
Db 2921 AGCCAGCGAGCAAGCTGCCCGCTGGGATCTCACAGATCAGCGCCCGCCACCTCTGCC 2980  
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900  
Db 2981 CCACAGCGCCCGCCCGCTCCCGCAGAAAGAGCTGCCCGGGGGCTGACAAGTCCACC 3040  
Qy 901 ProLeuThrAsnLysGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920  
Db 3041 CCATGTACCAACAAGGCGCAACCGAGAGGACTGTGGATCTCTCTGCNACCGAAGCTCTG 3100  
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940  
Db 3101 GGTCCTCTGCTCAATGCTATGCTCTGTCAGCGCCCTTCAGCCCATGCTTAGGAAGTCG 3160  
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960  
Db 3161 GCAACCAAGTTGAAGCGCTTAAGCGGGTAAAGCGCTCTATACTGTGTGGCTGACAACCCC 3220  
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980  
Db 3221 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGACGGGAGGAGGACCGAGAG 3280  
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000  
Db 3281 TGGTGGATTGSCACATTTGATGGAGATCTCTGTCGAAAGGCGCATTCGCCGTGTCTTT 3340  
Qy 1001 ValHisPheIleAlaAsp 1006  
Db 3341 GTGCACCTTTATCGCTGAC 3358



## RESULT 5

US-09-770-174-4023  
 ; Sequence 4023, Application US/09770174  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shyjan, Andrew W.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
 ; TITLE OF INVENTION: THEREFOR  
 ; FILE REFERENCE: 1600.2057-001  
 ; CURRENT APPLICATION NUMBER: US/09/770,174  
 ; CURRENT FILING DATE: 2001-01-26  
 ; PRIOR APPLICATION NUMBER: US 60/178,835  
 ; PRIOR FILING DATE: 2000-01-28  
 ; NUMBER OF SEQ ID NOS: 4376  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4023  
 ; LENGTH: 5756  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-770-174-4023

## Alignment Scores:

Pred. No.: 3,396-217 Length: 5756  
 Score: 5253.00 Matches: 1003  
 Percent Similarity: 99.70% Conservative: 0  
 Best Local Similarity: 99.70% Mismatches: 3  
 Query Match: 99.62% Indels: 0  
 DB: 34 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-770-174-4023 (1-5756)

Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAla	20
Db	334	ATGCCGGACCAAGATCTCCGTGCGGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG	393
Qy	21	ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	394	CCACGGCTTCAGCTTCACCAACCGCAGCGGAGTCCCGGAACACTGTGGCGCCATC	453
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrIlysMetLysSerValIysAla	60
Db	454	GAGGAGGCTTTGGACGTGACCGGATGCTTTTACAAATGAAGAAATCCGTGAAGCA	513
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnThrGlnAlaLeuGlu	80
Db	514	ATCAACAGCTCTGGCTGGCTCAGCTGGAATGAAGAGCAGTACACCCAGGCTCTGGAG	573
Qy	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuIys	100
Db	574	AAGTTTGGCGCACTGTGTATGCAGATGACCCAGATTTAGGAAGTGGCTTCTGAAG	633
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	634	TTCTCAGTGTACAAAGAGTTGACACACTTTTCAAAAACCTGATTCAGAAATATGAAC	693
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuIysGlyAspLeuLysGlyValIysGly	140
Db	694	AACATAATCTCTCTTCCCTTGGACAGTTTGTGTAAGGGGGACCTGAAGAGGAGTGAAGGG	753
Qy	141	AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrIysIle	160
Db	754	GATCTGAAAAAGCCCTTTTGATAAGCTTGGGAAGGACTATGAACAAAAATTAACCAAGATA	813
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	814	GAAAGGAGAAAAGGAACACGCCAAGCTCCATGGGATGATTCGACTGAATTAAGCGGA	873
Qy	181	AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	874	GCGGAAATTCGGAAGAGATGAAAGGAGAGGCGCTTCTTCCAGCTACAGATGTGCGAG	933
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	934	TATCTGCTGAAGGTCAACGAAATCAAGATTAAAAAGGGAGTAGATTTTACTTTTCAAGATCTG	993

Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	994	ATCAAAATACTTTTCATGCCCCAATGCAATTTTTCAGGATGGACTCAAAAGCCGTGAAAGC	1053
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1054	CTCAAACTTCCATGTAAACGCTGTCTACGGATCTTTCACAGATCAAAACAGGCCAGGAT	1113
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1114	GAAGAAAGAGCAGTGTGATACAGCTTCAGATATTTTGAATCCGCATTCGAGGTTGAA	1173
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1174	CAGAAAGAGGACTCCCAAAATTCGTACAGACAGCTTATAGCTTACATCAGCCTCAGGA	1233
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArg	320
Db	1234	AACAAGGAACATGGGACCGAGCGGAACGGAGCCCTCTACAAGAGAGTGACGGGATCCGA	1293
Qy	321	LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1294	AAAGTGTGGCAGAAAGGAATGTTTCAGTTAAAAATGTTTTCTGACCATATCCCATGGT	1353
Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1354	ACCGTAAACCGGCTCTCTGCAAGCTCAACCTGCTAACCTGCAGGTGAAGACCAACCT	1413
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1414	GAGGAGAAAGAGTCTTTGACCTTATTTTCATGACAGAACTTACCACTTTTCAAGCTGAA	1473
Qy	381	AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
Db	1474	GATGACAGGAATGTCAAATATGATGCTGTCTGTCGCAAAATAGCAAGAGAGAGCTTTA	1533
Qy	401	AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1534	AACAATGCTTTAAGGGGATGACAACTACTGGAGAAATAACATCGTCCAAGAACTGACA	1593
Qy	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1594	AAGAGAGATCACTCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGACTGTGGG	1653
Qy	441	AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1654	CGCGCAGATCTTACATGCTTTCCCAACCTGGGCATCTCTGACCTGATCGAGTGTTC	1713
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1714	GGAAATCCACGAGAGCTGGGGGTTTCATTTATCCAGGATGACGTCCCTGACCTTAGATGA	1773
Qy	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1774	CTGGGAACATCTGAGCTGCTGCTGCCAAGAAATATGGGAATGCAAGGCTTAATGAGATC	1833
Qy	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
Db	1834	ATGGAATGTTCCTTACCACTGAGGACTCAGTCAACCCCAACCCGCGGCGGACATGAT	1893
Qy	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysHis	540
Db	1894	GCAAGAAAGGACTACATCACAGCCCAAGTACATCCAGAGGAGATACGCAAGGAAGACAC	1953
Qy	541	AlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560
Db	1954	GCGGATAACGCGGGAAGCTTTCACAGCTTTTTCGAGGCGCGTCAAAACAGAGATATTTT	2013
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	2014	GGATGTCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCCAAC	2073

```
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2074 GGACATGACCGGATGAACCGCCCTCCACCTTTGCAGTCAGATCCGTGGATCGAACCTCT 2133
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2134 CTTCACATTGTAGACTTTTATTAGTCAGAACAGTGGGAACCTGGATATAACACAGCAGGAAA 2193
Qy 621 GlySerThrAlaLeuHisIleCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2194 GGCAGCAGCCCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2253
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaGlnSerGlyGluThrProLeuAspIle 660
Db 2254 CTGCGGGGAAGGCCCTCCATCGAGATAGCAACAGAGTCCAGAGAGACTCCCGTGGACATT 2313
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2314 GCCAAGCGCCCTCAAGCAGCAGCAGTGTGAGGAGCTGCTGACCCAGCCCTTATCTGGAAGA 2373
Qy 681 PheAsnSerHisValHisValGluThrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2374 TTTAATCTCAGTTTCACTGTTGAATATGAATGGCGACTACTCCACGAAGACCTGGATGAA 2433
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db 2434 AGTGATGACGACATGGATGAGAAATGTCAGCCCGAGTCCCAACCGCGGGAAGACCGGCC 2493
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2494 ATCAGCTTCTACCAGCTGGCTCCCAACAGCTTCACTAGCTTAACGCTGTATCTTTGGCCAGA 2553
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2554 GATGTCGAAACCTTGCCAAAGAGAAGCAGAGGGCTTTTCATGCCCAGCATCTTGCAGAAAT 2613
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2614 GAGACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGACGCCCCAGC 2673
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2674 ACCACGAGCCGCCCGCTCTCCACGGAATGTTGCCAAAGTTTCAGACAGCCTCTCT 2733
Qy 801 AlaAsnThrLeuThrLysThrAsnSerValSerValAspGlyLysArgGlnArgSer 820
Db 2734 GCTAACACCTGTGGAAGACAACTCTGTAAGTGTGGACGGTGGAAAGCCGCGAGCGATCT 2793
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2794 TCGTCAGATCCGCCAGCTGTCCATCCACGCTGCCCTCTCTTCGCGTGACATCTACCAAT 2853
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2854 CCCCTGACCCCGCCAGCCCGCTCCCGTGGCAAGAGCCCGCAGCGTAAATGGAAGCCCTG 2913
Qy 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2914 AGCCAGCCGAGCAACCTGCCCCGCTGGGATCTCACAGATCAGAGCCCGCCACCTCTGCC 2973
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2974 CCACAGCCCGCCAGCCCTCCCGCAGAAAGAGCCCTGCGCCGGGGGCTGACAAAGTCCACC 3033
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3034 CCACGTGACCAACAAAGGCCCAACCGAGAGAGACCTGTGGATCTCTCTGCAACGGAAGCTCTG 3093
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3094 GGTCTCTGTCCAATGCTATGCTGCTGCGCCCGCTGACCCACCTGACCCATGCTAGGAAGTCG 3153
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
```

```
Db 3154 GCACCAAGTTGAAGCTTACCGGGTGAAGCGCTCTATAACTGTGTGGCTGACACCC 3213
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3214 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGGACGGGAGGAGCCAGGAG 3273
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3274 TGGTGGATGGCCACATTGATGGAGATCCTGTGCGAAAGCGCATTTCCCGGTGTCAATT 3333
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3334 GTGCACCTTATCGCTGAC 3351
RESULT 6
US-60-172-360-28138
; Sequence 28138, Application US/60172360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Polynucleotide Sequence Polymorphisms Using
; FILE REFERENCE: GX-0007 P
; CURRENT APPLICATION NUMBER: US/60/172,360
; NUMBER OF SEQ ID NOS: 29838
; SOFTWARE: PERL Program
; LENGTH: 5720
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 331279.4
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 4927
; OTHER INFORMATION: a, t, c, g, or other
US-60-172-360-28138
Alignment Scores:
Pred. No.: 1,23e-216 Length: 5720
Score: 5240.00 Matches: 1003
Percent Similarity: 99.60% Conservative: 0
Best Local Similarity: 99.60% Mismatches: 3
Query Match: 99.37% Indels: 1
DB: 86 Gaps: 0
US-09-914-042-1 (1-1006) x US-60-172-360-28138 (1-5720)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCCGACCAAGATCTCCGTGTCGGAATTCGTGGCCGAGACCCATAGCAGCTACAAGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGCTCCAGCTTTCACACCGCCGAGTGCAGGAGACACCTGTGGCGGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGAGGGCTTTGGACGTGGACCGGATGGTCTTTTACAAATGAAGAAATCCGTGGAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCACGTGGAATAATGAAGACGAGTACACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGAGAGATGACCCAGATTTAGGAAGTGCCTTCTGAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
```

Db 641 TTTCTCAGTGTGTACAAAGAGTTGACGACCTTTTCAAAAAACCTCATTCAGAAATATGAAC 700  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTGAGAGGGGACCTCAAGAGGAGTGAAGGG 760  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 761 GATCTGAAAAAGCCCTTTTGATAAAGCTTGGAGGACTATGAACCAAAAAATAACCAAGATA 820  
Qy 161 GluLysGluLysLysGluHiAlaLysIleLeuHiSGLysMetIleAtrqThrGluIleSerGly 180  
Db 821 GAAAGAGGAAAAAGAAACACGCCAAGCTCCATGGGATGATTCGGACCTGAATAAGACGGA 880  
Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
Db 881 GCGGAAATTCGCGAAGAGATGAAAGAGGAGGCGCTTCTCCAGCTACAGATGTGCGAG 940  
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220  
Db 941 TATCTGCTGAAGGTCAACGAATCAAGATTAAGAGGAGTAGATTTACTTTCAGAAATCTG 1000  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 1001 ATCAAAATACTTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAAAGCCGTGGAAGC 1060  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1061 CTCAAACTTCCATGAAACGCTGTCTACGAGTCTTTCACAGATCAAAAGCCCGCAGAT 1120  
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1121 GAAGAAAGAGGAGCTTGATACAGCTTCGAGATATTTTGAATCCGATTCGAGGTGAA 1180  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 1181 CAGAAAGAGGACTCCCAATTCGTGAGAGCAGCTTATAGCTTACATCAGCCTCAGGGA 1240  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 1241 AACAAAGAACATGGAGCCGAGCGAGACGCTCTCAAGAAAGAGTACGCGATCCGA 1300  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle-SerHisG1 340  
Db 1301 AAAGTGTGGCAGAAAAAGAAATGTTTCAGTTAAATATGTTTCTGACCATATTCCTATGG 1360  
Qy 340 YThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPr 360  
Db 1361 TACCGCTAACCGGCCCTCTGCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCC 1420  
Qy 360 oGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaG1 380  
Db 1421 TGAGGAGAAAGAGTCTTTGACCTTATTTTCATGACAGAACTTACCACTTTCAAGCTGA 1480  
Qy 380 uAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLe 400  
Db 1481 AGATGAACAGGAATGTCAAAATATGATGCTGTGCTGCAAAATAGCAAGAAAGAGCTTT 1540  
Qy 400 uAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuTh 420  
Db 1541 AAACAATGCATTTAAGGGGGATGACAATACTGGAGAAAAATAACATCGTCCAAAGAACTGAC 1600  
Qy 420 rLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysG1 440  
Db 1601 AAAGAGATCATCTCAGAAAGTGACAGAGTACCGGGCAATGACGTCTGTGACTGTGG 1660  
Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460  
Db 1661 GCGGCAGATCTACATGGCTTTCACCAACTGGGCATCTGACCTGCATCGATCGAGTTC 1720  
Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVa 480

Db 1721 CGGAATCCACGAGAGCTGGGGGTTCTATTATTCAGGATGCAGTCCCTGCACCTTAGATGT 1780  
Qy 480 lLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 ACTGGGAACATCTGAGCTGTCTGCTCGCCAAGATATTTGGGAATGCAGGCTTTAATAGAT 1840  
Qy 500 eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs 520  
Db 1841 CATGGAATGTTGCCCTACCAGCTGAGGACTAGTCAAAACCAACCCAGGACGACATGAA 1900  
Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaAtrqLysLysHi 540  
Db 1901 TGCAAGAAGAGCTACATCACGCCAGATCATCGAGAGGAGATACGCAAGGAGAGCA 1960  
Qy 540 sAlaAspAsnAlaLysLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh 560  
Db 1961 GCGGATAAACGCGGAGGCTTTCAGTCTTTTGGAGGCCGTCAAAACGAGAGATATTTT 2020  
Qy 560 eGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs 580  
Db 2021 TGGATTTGCTCCAAAGCTTATGCTGATGTTGTGATCTTACGGAATAAATCCCACTGGCCAA 2080  
Qy 580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600  
Db 2081 CGGACATGAGCGCGATGAACGCGCCCTCCACTTGCAGTCAGATCCGTGGATCGAACCTC 2140  
Qy 600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy 620  
Db 2141 TCTTCACATTTGATAGCTTTTATGTTTCAGAACAGTGGGAACCTGGATTAACAGACAGGAA 2200  
Qy 620 sGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLe 640  
Db 2201 AGGACGACAGCCCTGCACACTACTGCTGCTGACCGACAATGCCAGTGCCTCAAGTTGCT 2260  
Qy 640 uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2261 CCTCGGGGGAAGGCTCCATCGAGATAGCAACAGAGTCAGAGAGACTCCCGCTGGACAT 2320  
Qy 660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 680  
Db 2321 TGCCNAGCGCTCAAGCAGAGACTGTGAGGAGCTGTGACCCCAAGCTTATCTGGAG 2380  
Qy 680 gPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspG1 700  
Db 2381 ATTTAAATTTCTCACGTTTCAGTTGAATATGAATGCGACTACTTCCACGAAGACCTGGATGA 2440  
Qy 700 uSerAspAspMetAspGluLysLeuGluProSerGluAsnArgArgGluAspArgPr 720  
Db 2441 AAGTGATGACGACATGGATGAGAAATTCGACGCCAGTCCCAACCGCGGGAAGACCGGCC 2500  
Qy 720 oIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaAr 740  
Db 2501 CATCAGCTTCTACAGCTGGGCTCCNACCAGCTTCAGTCTAAACGCTGTATCTTTGGCCAG 2560  
Qy 740 gAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760  
Db 2561 AGATGCTCAAAACCTTGCCAAGGAGAGACAGAGGCTTTTCATGCCAGCATCTTCAGAA 2620  
Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSe 780  
Db 2621 TGAGACTTTACGAGGCCCTCTCTGAGTGGCAGGCCACCTCCCGCCAGCTTCGACGCCCCAG 2680  
Qy 780 rThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSe 800  
Db 2681 CACCACAGGCCCCCCCGCTTCTCCACGGAATGTTGGCAAGTTTCAGACAGCCCTCTCTC 2740  
Qy 800 rAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSe 820  
Db 2741 TGCTAAACACCTCTGGGAAGACAACTCTGTAAGTGTGGACGCTGGAGCCCGCAGCGATC 2800  
Qy 820 rSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840  
Db 2801 TTCGTGATCCGCCAGCTGTCCATCCACCGCTGCCCTCTCTTCGCGTGACATCTACCAA 2860



380 uAspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAenSerLysGluGluAlaLe 400  
1462 AGATGAACAGGGAATGTCAAATATGATGTCTGTGTCGAAAATAGCAAGAAGAGCTTT 1521  
400 uAsnAenAlaPheLysGlyAspAsnThrClyGluAenAenIleValGlnGluLeuTh 420  
1522 AAACAATGCAATTTAAGGGGATGACAATCTGGAGAAATAACATCGTCCAAGAATGAC 1581  
420 rLysGluIleIleSerGluValGlnArgMetThrGlyAenAepValCysCysAspCysG1 440  
1582 AAAGGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGTCTGCTGTGACTGTGG 1641  
440 yAlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460  
1642 GCGCCAGATCTTACATGGCTTCCCAACACCTGGGCATCTCTGACCTGCATCGAGTGTTC 1701  
460 rClyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVa 480  
1702 CGGAATCCACCGAGAGCTGGGGGTTTATTTCCAGGATGCGAGTCCCTGAGCTTAGATGT 1761  
480 lLeuGlyThrSerGluLeuLeuAlaLysAenIleGlyAenAlaGlyPheAsnGluI1 500  
1762 ACTGGGAACATCTGAGCTGCTCGCCCAAGATATTTGGGAATGCGAGCTTTAATGAGAT 1821  
500 eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs 520  
1822 CATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCACCGCAGCGACATGAA 1881  
520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHi 540  
1882 TGCAAGAAGAGTACATACACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGCA 1941  
540 sAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh 560  
1942 CGCGGATTAACGGGGGAGCTTACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTT 2001  
560 eGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluTysIleProLeuAlaAs 580  
2002 TGGATTGCTCCAAAGCTTATGCTGTGTTGATCTTACGGAATAAAATCCCACTGCGCAA 2061  
580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600  
2062 CGGACATGAGCGGGATGAACCGCCCTCCACCTTCAGTCAGATCCGTGGATCGAACCTC 2121  
600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy 620  
2122 TCTTCACATTTGAGACTTTTATGTTACAGACAGTGGGAACCTGGATTAACAGACAGGAA 2181  
620 sGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLe 640  
2182 AGGCAGCACAGCCCTGCACCTACTGCTGCTGACCGACAATGCGAGTGCCTCAAGTTGCT 2241  
640 uLeuArgLysAlaSerIleGluIleAlaAenGluSerGlyGluThrProLeuAspI1 660  
2242 CCTCGGGGAGGCGCTCATCGAGATAGCAACGAGTCAGGAGAGACTCCGCTCGACAT 2301  
660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 680  
2302 TGCCGAGCGCTCAAGCAGAGACTGTGAGAGAGCTGTGACCCCAAGCTTTATCTGGAAG 2361  
680 gPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspG1 700  
2362 ATTTAATTCACGTTACGTTGATATATGATGATGAGTACTCTCCACGAAGACCTTGATGA 2421  
700 uSerAspAspAspMetAspGluLysLeuGlnProSerGluAenArgArgGluAspArgPr 720  
2422 AAGTATGACGACATGGATGAGAAATTCAGCCCACTGCCAACCCGCGGGAAGACCGGCC 2481  
720 oIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAenAlaValSerLeuAlaAr 740  
2482 CATCAGCTTCTACCAAGCTGGGCTCCAAACAGCTTACGTTACAGCTGTATCTTTGGCCAG 2541

740 gAspAlaAlaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760  
2542 AGATGCTCAAAACCTTGCAGAGGAGAGCAGAGGGCTTTTCATGCCAGCATCTTCAGAAA 2601  
760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSe 780  
2602 TGAGACTTACGAGGCGCTTCTGAGTGGCAGCCACCTCCGCGCCAGCTTGCAGGCCCCAG 2661  
780 rThrThrSerAlaProProLeuProProArAenValGlyLysValGlnThrAlaSerSe 800  
2662 CACCACAGCGCCCGCCCTCTCTCCACGGAATGTTGGCAAGATTTCAGACAGCTCTCTC 2721  
800 rAlaAenThrLeuTyrLysThrAenSerValSerValAspGlyGlySerArgGlnArgSe 820  
2722 TGCTAACACCTCTGGAAAGACAAACTCTGTAAAGTGGACGGTGAAGCCGCGCAGCATC 2781  
820 rSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840  
2782 TTGCTCAGATCCGCGAGCTGTCCATCCACCGCTGCCCTCTTTCGCGTGACATCTACCAA 2841  
840 nProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLe 860  
2842 TCCCTGACCCCGCCCGCCCGCCCGTTCGCAAGACGCCACCGTAAATGGAAGCTT 2901  
860 uSerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPr 880  
2902 GAGCCAGCGAGCAGCGCTGCGCCGCTGGGATCTTCACAGATCAGCGCCCGCCACCTCTGCC 2961  
880 oProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900  
2962 CCCAGCGCCCGCCCGCCCGCTCCGCGAAGAAGCGCTCGCGCGGGGCTGACAAGTCCAC 3021  
900 rProLeuThrAsnLysGlyGlnProArArgGlyProValAspLeuSerAlaThrGluAlaLe 920  
3022 CCCACTGACCAACAAAGCCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 3081  
920 uGlyProLeuSerAenAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940  
3082 GGGTCTCTGTCCAATGCTATGGTCTCTGACGCGCCCTGCACCATGCTTAGGAAGTCGCA 3141  
940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAenCysValAlaAspAenPr 960  
3142 GGCAACCAAGTTGAAGCCTGAGCGGGTGAAGCGCTCTATACTGTGTGGCTGACAACCC 3201  
960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980  
3202 CGATGAGCTCACCCTTCTCCGAGGGGATGTGATCATCTGTCGCGGGAGGAGGACAGGA 3261  
980 uTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
3262 GTGGTGGATTGGCCACATTTGATGGAGATCTCTGGTCGCAAGCGCATTTCCCGGTGTCAAT 3321  
1000 eValHisPheIleAlaAsp 1006  
3322 TGTGACCTTTATCGCTGAC 3340

RESULT 8  
US-60-278-232-6642  
; Sequence 6642, Application US/60278232  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide  
; FILE REFERENCE: GX-0011 P  
; CURRENT APPLICATION NUMBER: US/60/278,232  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 12,557  
; SOFTWARE: PERL Program  
; SEQ ID NO 6642  
; LENGTH: 5810

[illegible]

```
Db 2242 CTGCGGGGAAGGCTTCATCGAGATAGCAACGAGTCAGGAGAGACTCCGCTGGACAT 2301
Qy 660 ealalyArGleuLysHisGluHisCysGluGluLeuThrGlnAlaLeuSerGlyAr 680
Db 2302 TGCCAAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGTGACCAAGCCCTTATCTGGAAG 2361
Qy 680 gheAenSerHisValHisValGluTyroGluTrpArgLeuLeuHisGluAspLeuApG1 700
Db 2362 ATTAAATTCACGTTACGTTGAATATGAATGGGAGCTACTCCACGAAGACCTGGATGA 2421
Qy 700 uSerAspAspMetAspGluLysLeuGlnProSerGluAenArgArgGluAspArgPr 720
Db 2422 AAGTGATGACGATGGATGGAAATTTGAGCCAGTCCCAACCGCGCGGAAGACCGGCC 2481
Qy 720 oileSerPheTyroGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaAr 740
Db 2482 CATCAGCTTCTACGAGTGGGCTCCCAACGAGCTTCAGTCTAAGCGTGATCTTTGGCCAG 2541
Qy 740 gAspAlaAlaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760
Db 2542 AGATGCTGCAACCTTGCAGAGGAAGCAGAGGGCTTTTCATGCCCGCAGCATCTTGCAAA 2601
Qy 760 nGluThrTyroGlyAlaLeuLeuSerClySerProProAlaGlnProAlaAlaProSe 780
Db 2602 TGAGACTTACGGAGCGCTCTCAGTGGCAGCCACCTCCCGCCAGCCTGCAGCGCCCGCAG 2661
Qy 780 rThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSe 800
Db 2662 CACCACAGCGCCCCCGCTCTCTCCAGGAATGTTGCCAAGTTTCAGACAGCTCTCTC 2721
Qy 800 rAlaAenThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSe 820
Db 2722 TGCTAACACCTGTGGAGACAACTCTGTAAGTGTGGACGCTGGAAGCGCGCAGCGATC 2781
Qy 820 rSerSerAspProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840
Db 2782 TTCGTGACATCGCCAGCTGTCCATCCCGCTGCGCCCTCTCTCGGTGACATCTACCAG 2841
Qy 840 nProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLe 860
Db 2842 TCCCTGTACCCCGCAGCGCCCGCTCCAGCGGTTGCCAAGACGCCCGCGTAAAGAGCCTT 2901
Qy 860 uSerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPr 880
Db 2902 GAGCAGCGCAGCAAGCCTGCGCCCGCTGGGATCTCACAGATCAGCGCCCGCCACCTCTGCC 2961
Qy 880 oProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900
Db 2962 CCCACAGCGCCCGCAGCGCCCTCCCGCAGAAGAAGCTGCGCGGGGGCTGACAAGTCCAC 3021
Qy 900 rProLeuThrAenLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920
Db 3022 CCCACTGACCAACAAAGGCCAACCGAGAGACCTGTGGATCTCTCTGCAACGGAAGCTCT 3081
Qy 920 uGlyProLeuSerAenAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 3082 GGGTCTCTGTCCATGTATGTCTCTGAGCGCCCTGACCCCATGCTAGGAAGTGCAG 3141
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyroAenCysValAlaAspAsnPr 960
Db 3142 GGCACCAAGTTGAAGCGCTGAGCGGGTGAAGCGCTCTATTAAGTGTGGCTGACAACCC 3201
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
Db 3202 CATAGAGCTACCTTCTCCGAGGGGATGTGATCATCGTGGACGGGGAGGAGCAGGA 3261
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 3262 GTGGTGGATGGCCACATTTGATGAGATCCTGTGTCGAAGGCGCATTTCCCGGTCTATT 3321
Qy 1000 eValHisPheIleAlaAsp 1006
|||||
```

```
Db 3322 TGTGCACTTTATCGCTGAC 3340
RESULT 9
US-09-572-411-6530
; Sequence 6530, Application US/09572411
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Xu, Yong Yao
; APPLICANT: Acton, Susan L.
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USBS
; FILE REFERENCE: 1600.1091-001
; CURRENT APPLICATION NUMBER: US/09/572,411
; EARLIER FILING DATE: 2000-05-12
; EARLIER APPLICATION NUMBER: 60/133,993
; EARLIER FILING DATE: 1999-05-13
; EARLIER APPLICATION NUMBER: 60/135,617
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,816
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,817
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,950
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,679
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,630
; EARLIER FILING DATE: 1999-05-24
; NUMBER OF SEQ ID NOS: 7417
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6530
; LENGTH: 6389
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(6389)
; OTHER INFORMATION: n = A,T,C or G
US-09-572-411-6530

Alignment Scores:
Pred. No.: 2,7e-215 Length: 6389
Score: 5210.00 Matches: 996
Percent Similarity: 99.11% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 9
Query Match: 98.81% Indels: 0
DB: 26 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-572-411-6530 (1-6389)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyroLysAla 20
|||||
Db 341 ATCGCGGACCAAGATCTCCGTCGGAATTCGTGGCCGAGACCCATGAGACTACAAAGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
|||||
Db 401 CCCACGGCTCTCAGCTTCACCACCGCGCAGTCGCGGACACACACTGTGGCGGCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyroLysMetLysLysSerValLysAla 60
|||||
Db 461 GAGGAGGCTTTGGAGCTGGACCGGATGTTCTTTACAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyroThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGCGAAATGAAGAGCAGTACACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAenCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
|||||
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTGCCTTCTTGAG 640
```



Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db 641 TTCTCAGTGTGTTTACAAAGGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAAATATGAAC 700  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTGAAGGGGGACCTGAAAGGAGTGAAGGG 760  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 761 GATCTGAAAAAGCCCTTTTGATTAAGCTTGGAAAGGACTATGAACAAAAAATAACCAAGATA 820  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysIleSerGly 180  
Db 821 GAAAAGGAAAAGGAACACGCCNAGCTCCATGGGATGATTCGGACTGAATAATAGCCGA 880  
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGlu 200  
Db 881 GCGGAAATGTCGGAAGAGATGGAAGAGGAGAGGGCGCTTCTCCAGCTACAGATGTGCGAG 940  
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 941 TATCTGCTGAAGGTCAACGAAATCAAGATTAAAAAGGGAGTAGATTTTACTTTCAGAAATCTG 1000  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 1001 ATCAAAATACTTTTCATGCCCAATGCNAATTTTTTTCAGGATGGACTCAAAAGCCGTGGAAAGC 1060  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1061 CTCAAACTTCATTTGAAACGCTGTCTACGGATCTTCACAGATCAAAACAGCGCCAGGAT 1120  
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1121 GAAGAAAGAGCGCAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCAGGTTGAA 1180  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 1241 AACAAAGAAATGCGACCGAGCGGAACGGCAGCTCTCAAGAGAGTGACGGGATCGGA 1300  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1301 AAAGTGTGGCAGAAAAGGAAATGTTTCAGTTAAAAATGGTTTTCTGACCATATCCCATGGT 1360  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1361 ACCGCTAACCGGCTCTCTGCAAGCTCAACCTGTACCTGCCAGTGGAGAACCAACCTT 1420  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1421 GAGGAGAAGAAGTGTCTTGACCTTATTTACATGACAGAACTTACCACTTTCAAGCTGAA 1480  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAAATAGCAAAAGAAAGCTTTA 1540  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1541 AACAAATGCAATTAAGGGGATGACAAATCTGGAGAAATAATCATCTGCTCAAGAACTGACA 1600  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1601 AAGGAGATCATCTCAGAAGTCAGAGGATGACGGCAATGACGTCTGTGTGACTGTGGG 1660  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1661 GCGCAGATCTCATACATGCTTTCCACCAACTGGGCATCTCGACCTGCATCGAGTGTTC 1720  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480

Db 1721 GGAATCCACCGAGAGCTGGGGGTTTCATATTTCAGGATGCAGTCCCTGACCTTAGATGTA 1780  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 CTGGGAAACATCTGAGCTGCTGCTGCCAAGAATATTGGGAATGCAGGCTTTAAATGAGATC 1840  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1841 ATGGAATGTTCCTTACCAGCTGAGAGCTCAGTCAAAACCCAAACCCAGGCGGACATGANT 1900  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1901 GCAAGAAAGGACTACATCACAGCCAGTACATCCAGAGGAGATACGCAAGGAAGAGCAC 1960  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1961 GCGGATAACGCGGCGAAGCTTCACAGTCTTTTGCAGGCGGTCAAAACAGAGAGATATTTTT 2020  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2021 GGATGTCTCCAAAGCTTATGCTGATGGTGGATCTTACGGAAATAATCCCCACTGGCCAAC 2080  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTCGAGTCAGATCCGTGGATCGAACCTCT 2140  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2141 CTTTCACATTTAGACTTTTTTAGTTTCAGAACAGTGGAAACCTGGATAAAACAGACAGGAAA 2200  
Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2201 GGCAGCAGAGCCCTGCACTACTGCTGCTGACCCACAATGCCGAGTGGCTCAAGTTGCTC 2260  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2261 CTGCGGGGGAAGGCTCCATCGAGATAGCAAAACAGTCAGGAGAGACTCCGCTGGACATT 2320  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2321 GCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGTGACCCAAGCCTTATCTGGAAGA 2380  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2381 TTTAAATCTCAGCTTACGTTGAATATGAATGGGAGTACTTCCACGAAGAGCTGGATGAA 2440  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2441 AGTCATGACGACATGGATGAGAAATTCAGCCCACTCCCAACCGCGGGAAGACCGGCC 2500  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2501 ATCAGCTTACAGCTGGGCTCCAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560  
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2561 GATGCTGCAAAACCTTGCCAAGGAGAGCAGAGGCTTTTCATGCCAGCATCTTGCAGAA 2620  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2621 GAGACTTACGGAGCCCTCCTGAGTGGCAGCCCACTCCCGCCAGCTGCAGCCCCCAGC 2680  
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2681 ACCACCGCGCCCCCGCTTCTCCACGGAATGTTGGCAAAAGTTTCAGACAGCTCTCTCT 2740  
Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
Db 2741 GCTAACACCTTGTGGAGACAAACTCTGTAAAGTGGCTCGCTGTAGCGCGCGCGGCCCCC 2800  
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840  
::: |||||



```
Db 2801 CGGTGAGATCGCCAGCTGTGATCCACCGCTGCCCTCTTTCGGGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCACACGCGCCGCCCTGTCAGAGACGCCAGGTAATGGAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysGlnLysArgProProProProPro 880
Db 2921 AGCCAGCGAGCAAGCCGCGCTGCGGATCTCACAGATCAGGCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCGCCGCGCTCCGCGAGAGAGCCCTGCGCGGGGCTGACAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTGTCAACGGAGCTGTG 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGCTGCTGCGAGCCCTGCAACCCCTGCAAGTCCGAG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuThrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValLysLysValLysLysGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTCTCCAGGGGATGTGATCATCTGTGGAGCGGGAGGAGCAGGAG 3280
Qy 981 TrpThrLysGlyHisLysAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATGGCCACATGATGAGATCTGTGTGCGAAAGCGCATTCCTCCGCTGTCAAT 3340
Qy 1001 ValHisPheLysAlaAsp 1006
Db 3341 GTGCACITTTTCGCTGAC 3358

RESULT 10
US-10-170-235-8470
; Sequence 8470, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; FILE OF INVENTION: TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
; FILE REFERENCE: CL001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 8470
; LENGTH: 5670
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-8470

Alignment Scores:
Pred. No.: 7,78e-213 Length: 5670
Score: 5152.00 Matches: 988
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 2
Query Match: 97.71% Indels: 16
DB: 49 Gaps: 1

US-09-914-042-1 (1-1006) x US-10-170-235-8470 (1-5670)

Qy 1 MetProAspGlnLysSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCGGACCAAGATCTCGGTGCGGAATTCGTGGCCGAGACCCATGAGGACTACAAGCGC 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaLys 40
Db 401 CCCAGCGCTTCAGCTTCACCACCGCACGCGGAGTCCCGGAACACTGTGGCGGCCATC 460
```

```
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACGTGCGACCGATGCTTTTACAAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValClnAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGGAAATGAAGAGCAGTACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTCCGTCTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGGAGCTTCACAGCACTTTTCAAAAACCTGTATTCAGATATGAAC 700
Qy 121 AsnLysLysSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTGAAGGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysLysLysLys 160
Db 761 GATCTGAAAAGCCTTTTGATAAAGCTTGGAGGAGCTATGAAACAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetLysLysLysLysLysLys 180
Db 821 GAAAGGAGAAAAGGAAACAGCCAGCTCCATGGATGATTCGGACTGAAATAAGCGGA 880
Qy 181 AlaGluLysLysGluLysGluMetGluLysGluLysGluLysGluLysGluLysGlu 200
Db 881 GCGGAAATTTGCGAAGAGATGGAAGAGAGAGGCGCTTCTTCCAGCTACAGATGTCGAG 940
Qy 201 TyrLeuLysLysValAsnGluLysLysLysLysLysLysLysLysLysLysLysLys 220
Db 941 TATCTGCTGAAGTCAACGAATCAAGATTTAAAAGGAGTAGATTTACTTTCAGATCTG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAATACTTTTCATGCCAATGCAATTTTTTTCAGGATGGACTCAAAAGCGTGAAGAG 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrLysGluAlaGlnAsp 260
Db 1061 CTCAAACCTTCATTTGAAACGCTGCTACGAGTCTTACACAGATCAAAACAGGCCAGAT 1120
Qy 261 GluGluArgArgGlnLeuLysGlnLeuArgAspLysLysLysLysLysLysLysLys 280
Db 1121 GAAGAAAGAGGAGGAGTGTATACAGCTTCGAGATATTTTGAATCCGCATTCGAGGTTGA 1180
Qy 281 GlnLysGluAspSerGlnLysArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGA 1240
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLysArg 320
Db 1241 AACAAAGGAACATGGAGCGGAGCGGAGCGGAGCGCTCTACAAAGAGAGTGAAGGAGTCCGA 1300
Qy 321 LysValThrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrLysSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAAGGAATGTTCAAGTTAAAATGGTTTTCTGACCATATCCCATGGT 1360
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGCGCTCTCTGCAAAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCT 1420
Qy 361 GluGluLysLysCysPheAspLeuLysSerHisAspArgThrThrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAGAGAGTCTTTGACCTTATTTTCAATGACAGAACTTACCACCTTTCAAGCTGAA 1480
Qy 381 AspGluGlnGluCysGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 400
Db 1481 GATGAACAGGAGATGTCATATGATGCTGTGTCGCAAAAATGCAAAAGAGAGAGCTTTA 1540
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
```

```
Db 1541 AACATGATTTAAGGGGAGTACATCTGGAGAAATAACATCGTCCAAAGACTGACA 1600
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCGTGTGACTGTGGG 1660
Qy 441 AlaProAspProThrTrrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCAGATCCATACATGGCTTTCCACCAACCTGGGCATCTCCTGACCTGATGGTGTCC 1720
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1721 GGATTCACAGAGAGCTGGGGTTCATTTATCCAGGATGCACTCCTGACCTTAGATGTA 1780
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGCAACATCTGAGCTGCTGCTGCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 1840
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTCCCTACCAAGCTGAGGACTCAGTCAAAACCCCAAGGAGGACATGAAT 1900
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTACATCAGAGCCAGTACATCGAGAGAGATACGCAAGGAAGACAC 1960
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 GCGGATAACGCGCGAAGCTTCACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTT 2020
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATGTGCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAA 2080
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTGCACTGAGTCAGATCCGTGGATCGAACCTCT 2140
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2141 CTTCACTGTGAGACTTTTATGTTTACAGAACAGTGGGAACTTGATATAACACAGACAGGAAA 2200
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2201 GGCAGCAGACCGCTGCTACTGCTGCTGCTGCCAGCAATGCGAGTGCCTCAAGTTGCTC 2260
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGAGAGCCCTCCATCGAGATAGCAACAGAGTCAGAGAGACTCCCGTGGACATTT 2320
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAACGACGAGCAGCTGTGAGGAGCTGCTGACCCAAAGCCTTATCTGGAAGA 2380
Qy 681 PheAsnSerHisValHisValGluTyrGluTrrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAATCTCAGTTTACGTTTGAATGAATGCGCACTACTCCACGAAGACCTGGAATGAA 2440
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db 2441 AGTGAATGACATGGATGAGAAATGTCAGCCAGTCCCAACCGCGGGAGACCGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACCACTGGGCTCCAAACAGCTTCAGTCTAAGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAAAACCTTGGCAAGGAGAGCAGAGGGCTTTTCATGCCAGCATCTTTGAGAAAT 2620
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
```

```
Db 2621 GAGACTTTACGAGGACCCCTCTGAGTGGCAGCCCACTCCCGCCAGCCTGCAGCCCCCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCCCGCTTCTCCACGGAATGTTGGCAAAGTTCAGACAGCCTCTCTCT 2740
Qy 801 AlaAsnThrLeuTrrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCCCTGTGGAGAGCAAACTCTGTAAAGTGTGGACGGTGAAGCCGCGGATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCCAGCTGTCATCCACGCTGCCCTCTCTTCGGGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCAACGCGCCGCCCAACCGCTTGGCAAGACGCCAGCGTAATGGAAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2921 AGCAGCGGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCCACTCTGCCCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCAACGCGCTCCCGCAGAGAAGCCTGCGCCGGGA----- 3025
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3026 -----GTGGATCTCTCTGCAACGGAAGCTCTG 3052
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3053 GGTCTCTGTCCAATGCTATGCTCTGAGCCCTCGCAGCCCTGACCATGCTTAGGAAGTCGAG 3112
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3113 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATAACTGTGTGGCTGACAACCCC 3172
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3173 GATGAGCTCACTTCTCCGAGGGGATGTGATCATCTGTGGACGGGGAGGAGGACGAGAG 3232
Qy 981 TrrPrrIleGlyHisIleAspGlyAspProGlyArgLysGlyValAlaPheProValSerPhe 1000
Db 3233 TGGTGGATGTCACCATTTGATGGAGATCTCTGTCGAAAGCGCATTTCCCGGTGTCTATT 3292
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3293 GTGCACCTTTATCGCTGAC 3310
```

## RESULT 11

```
US-60-452-680-5751
; Sequence 5751, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5751
; LENGTH: 5670
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-5751

Alignment Scores: 1.41e-212 Length: 5670
Pred. No.: 5146.00 Matches: 987
Score: 98.11%
Percent Similarity: 98.11% Conservative: 0
```

Best Local Similarity: 98.11% Mismatches: 3  
Query Match: 97.59% Indels: 16  
DB: 115 Gaps: 1

US-09-914-042-1 (1-1006) x US-60-452-680-5751 (1-5670)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
Db 341 ATGCCGACCAAGATCTCGGTGCGAATTCGTGGCCGAGACCCTACGAGGCG 400

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
Db 401 CCCAGGCTCCAGCTTCCACCCCGCAGCGCCAGTCCCGGACACACATGTGGCGCCATC 460

Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
Db 461 GAGGAGGCTTTGGAYGTGGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80  
Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACACCCAGGCTCTGGAG 580

Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
Db 581 AAGTTTGGCGGCACTGTGTATGCAGATGACCCAGATTTAGGAAGTGGCTTCTGAAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAGATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 701 AACATAATCTCTCTCCCTTTGACAGTGTGCTGAAGGGGGAGCTCAAAAGAGTGAAGGG 760

Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrLysIle 160  
Db 761 GATCTGAAGAAAGCCCTTTTGATAAGCTTGGAGGAGCTATGAACAAAATTAACCAAGATA 820

Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180  
Db 821 GAAAGGAGAGAAAAGGACACGCCAAGCTCCATGGGATGATTCGGACTGAAATGAAGCGA 880

Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
Db 881 GCGGAAATTTGCCAGAGATGCGAAAGGAGAGGCGCTTCTCCAGCTACAGATGTGGAG 940

Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220  
Db 941 TATCTGCTGAAGGTCACGAAATCAAGATTAAGAGGGAGTAGATTTACTTCAGAAATCTG 1000

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 1001 ATCAAAATCTTCATGCCCAATGCAATTTTTCAGGATGACTCAAGCGCTGGAAAGC 1060

Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1061 CTCAAACCTTCATTTGAACGCTGTCTACGGATCTTCACAGATCAAAACAGCGCCAGGAT 1120

Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1121 GAAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCAGGTTGAA 1180

Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 1181 CAGAAAGAGGACTCCCAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240

Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 1241 AACAGGAAACATGGGACCGAGCGGAACCGAGCCTCTACAAAGAGAGTGAAGGATCCGA 1300

Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1301 AAAGTGTGGCAGAAAGGAAATGTTTCAGTTAAATATGTTTCTGACCATATATCCATGGT 1360

Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1361 ACCGCTAACCGGCTCTCTGCAAGGCTCAACTGCTTAACCTGCAGTGAAGACCAACCTCT 1420

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1421 GAGGAGAAAGAGTGTCTTTGACCTTATTTTCATCATGACAGAACTTACCACCTTTCAAGCTGAA 1480

Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1481 GATGAAACAGGAATGTCAAAATATGATGTCTGTGTCGCAAAATAGCAAGAGAAAGACTTTA 1540

Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1541 AACAAATGCAATTAAGGGGATGACAAATCTCGAGAAATAACATCGTCCAGAACTGACA 1600

Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGAYGTCTGTGTGACTGTGGG 1660

Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1661 GCGCAGATCTCTACATGGCTTTCCACCAACTGGGCATCTTGACCTGCATCGAGTGTCC 1720

Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1721 GGAATCCACCCAGAGAGTGGGGTTCATTTATTCAGGATGCAGTCTCGACCTTAGATGTA 1780

Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1781 CTGGGAACATCTGAGCTGCTCGCCAAAGAAATATTTGGGAATGCAAGGCTTTAATGAGATC 1840

Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1841 ATGGAATGTTCCCTTACCAGCTGAGACTCAGTCAAAACCCAGGAGGAGCGCATGAT 1900

Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1901 GCAAGAAAGGACTACATCACAGCCAAAGTATATTCAGAGAGAGATATCGCAAGGAAGACAC 1960

Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1961 GCGGATACGCGGCGAAGCTTTCAGAGTCTTTGCGAGGCGCTCAAAACAGAGAGATATTTT 2020

Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 2021 GGATTTCTCCAAAGCTTATGCTGATGGTGGATCTTACGGAATAAATCCCTGCTGCCAAC 2080

Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTTCGCTGATCGATCCGTGATCGAACCTCT 2140

Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2141 CTTCACATTTGATGACTTTTATAGTTTCAGAAACAGTGGAACTTGGATTAACAGACAGGAAA 2200

Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2201 GGCAGACAGCCCTGCACTACTGCTGCTGCCCAATGCGAGTGGCTCAAGTTGCTC 2260

Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660  
Db 2261 CTGCGGGGAGGCTCCATCGATGACAAAGAGTCAAGTCAAGAGAGACTCCGCTGGACATT 2320

Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2321 GCCAAGGCGCTCAAGACAGCAGCACTGTGAGGAGCTGTGACCCAGCCTTATCTGGAAGA 2380

Qy 681 PheAsnSerHisValHisValGluTyrClnTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2381 TTTAATTTCTCAGTTTCAAGTTGAATATGATGGGAGCTACTTCCAGGAAGACCTGGATGAA 2440

```
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTGATGACGACATGGATGAGAAATTGCGCCAGTCCCAACCGCGGGAAGACGGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACCAAGTGGCTCCCAACCAAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAAACTTTGCCAAGASAAAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAA 2620
Qy 761 GluThrTyrGlyAlaLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGAGAGCCCTCTGAGTGGAGCCACCTCCCGCCAGCCTGCGAGCCCCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCGCCGCTTCTCCACGGAATGTTGGCAAAAGTTCAGACAGCCTCCTCT 2740
Qy 801 AlaAsnThrLeuTyrPheThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTGTGGAAGACAACTCTGTAAAGTGTGGACGGTGGGAAGCGCGACGCATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCACAGCTGTCTCATCCACGCTGCCCTCTTTCGGGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCCCGCCCGCCCGCTGCCAAGACGCCAGCGTAATGGAGCCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGCAGCAAGCCCTGCCCGCTGGGATCTCACAGATCAGCGCCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCGCGCTCCCGCAGAAAGCCCTGCCCGCGGA----- 3025
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3026 -----GTGGATCTCTCTGCAACGGAAGCTCTG 3052
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3053 GGTCCTCTGTCCAATGTCTATGCTCTCGACGCCCGCTGCACCCCATGCCCTAGGAAGTCGCAG 3112
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3113 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCCC 3172
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3173 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCGTGGAGCGGGAGGAGCACCAGGAG 3232
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3233 TGGTGGATTGGCCACATTGATGGAGATCTGTGTCGAAAGGCGCATTCCTCCGCTGTCAVTT 3292
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3293 GTGCACCTTATYGCTGAC 3310
```

## RESULT 12

```
US-09-949-002-139
; Sequence 139, Application US/09949002
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; AND USES THEREOF
; FILE REFERENCE: CL000790
```

```
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 139
; LENGTH: 3025
; TYPE: DNA
; ORGANISM: Human
; US-09-949-002-139

Alignment Scores:
Pred. No.: 5,44e-192 Length: 3025
Score: 4664.00 Matches: 893
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 2
Query Match: 88.45% Indels: 0
DB: 41 Gaps: 0
```

US-09-914-042-1 (1-1006) x US-09-949-002-139 (1-3025)

```
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCCGACCCAGATCTCCGTGTCGGAATTCGTGCCGAGACCCATGAGGACTCAAGGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGCGCTCCAGCTTTCACCCACCGCAGCGCGAGTCCCGGAACACTGTGGCGCGCATC 460
Qy 41 GluGluAlaLeuAspValAspAtqMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACGTGGACCGGATGCTTTTACAAATGAAGAAATCCGTGGAAGCA 520
Qy 61 IleAsnSerSerGlyLysAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCACGTGGAATGAAGAGCAGTACACCCAGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTGCCTTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTCAAAGAGGTTCAGACACTTTTCAAAACCTGATTCAGATATGNAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTTCCCTTTGGACAGTTTGTGAAGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 761 GATCTGAAAAAGCCCTTTTGTATAAAGCTTGGAAAGACTATGAAACAAAAAATACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAGAGAAAAAGAACACCCACAGCTCCATGGGATGATTCGAGCTGAATAAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGGAAGAGGAGGCGCTTCTTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLysLysValAsnGluIleLysIleLysLysGlyValAspLeuGluAsnLeu 220
Db 941 TATCTGCTGAAGGTCAACGAAATCAAGATTTAAAAGGGAGTAGATTTACTTCAGAACTG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValIleSer 240
Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTTTCAGGATGCACTCAAGCGCTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTTCCATTGAAACGCTGTCTACGGATCTTTCACAGCATCAAAACAGGCCCGAGAT 1120
```

QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
DB 1121 GAAGAAAGAGGACGTTCGATACAGCTTCGAGATATTTTGAATCCGATTCGAGTTGAA 1180  
QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
DB 1181 CAGAAAGAGGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTATACATCAGCCTCAGGGA 1240  
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
DB 1241 AACAGGAACATGGGACCGGAGCGGAAACGCGAGCCTCTACAGAAAGAGTGACGGGATCCGA 1300  
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
DB 1301 AAAGTGTGCGCAGAAAGGAAATGTTTCAGTTAAATAATGGTTTCTGACCATATCCCATGGT 1360  
QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
DB 1361 ACCGCTAACCGGCTCCCTGCAAGAGCTCAACCTGCTTAACCTGCGAGTGAAGACCAACCTT 1420  
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
DB 1421 GAGGAGAAGAGTGTCTTGACCTTATTTCAATGACAGACTTACCACTTTCAAGCTGAA 1480  
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
DB 1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAGAAAGAGCTTTA 1540  
QY 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnThrGlyGlnGluLeuThr 420  
DB 1541 AACATGTCATTTAAGGGGAGTGAACAATCTCGAGAAATAATCATCTGTCNAAGAACTGACA 1600  
QY 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
DB 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGTGACTGTGGG 1660  
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
DB 1661 CGCCAGATCTACATGCTTTCACCAACCTGGGCATCTGACCTGCATCGAGTGTTC 1720  
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
DB 1721 GGAATCCACCGAGAGCTGGGGGTTCAATTTATTCAGGATGCACTCCCTGACCTTAGATGA 1780  
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
DB 1781 CTGGGAACATCTGAGCTGCTGCTGCCAAGAAATTTGGGAATGCAAGGCTTTAATGAGATC 1840  
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
DB 1841 ATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGCGAGCATGAAT 1900  
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
DB 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGACAC 1960  
QY 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
DB 1961 CGGATACCGCGGAGACCTTCACAGTCTTTGCGAGGCGGTCAAAACCGAGAGATATTTT 2020  
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
DB 2021 GGATTTCTCCAAAGCTTATGCTGATGGTGTGGATCTTACGGAAAAATCCCACTGGCCAA 2080  
QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
DB 2081 GGACATGAGCGGATGAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140  
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrClyLys 620  
DB 2141 CTTCACTTGTAGACTTTTATAGTTCAGAAACAGTGGGAACCTGGATAAACACAGACAGGGA 2200

QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
DB 2201 GGCAGCAGAGCCCTGCACTACTGTGCTGACCCAGCAATGCGAGTGCCTCAAGTTGCTC 2260  
QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
DB 2261 CTGCGGGGAAGGCTTCATCAGATAGCAACGAGTCAGGAGACTCCCGCTGGACATT 2320  
QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
DB 2321 GCCAAGCGCCTCAGCAGCAGCAGCTGTGAGGAGCTGTGACCCAAAGCTTATCTGGAAGA 2380  
QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
DB 2381 TTTAATTTCTACGTTTACGTTGAATATGAATGGGACTACTCCAGAAAGACCTGGATGAA 2440  
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
DB 2441 AGTGAATGACGATGGATGAGAAATTTGAGGCCCTCCAAACCGCGGGAAGACCGGCC 2500  
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
DB 2501 ATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCAGTCTAAGCTGTATCTTTGGCCAGA 2560  
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
DB 2561 GATGCTGCAAACTTGCAGGAGAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAAAT 2620  
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780  
DB 2621 GAGACTTACGAGAGCCCTCCTGAGTGGCAGCCCACTCCCGCCCGAGCTGCAGCCCGCAGC 2680  
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800  
DB 2681 ACCACCGAGCCCGCCCGCTTCTCCACGGAATGTTGGCAAGTTTCACAGCCTCTCTCT 2740  
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
DB 2741 GCTAACACCTGTGGAAGACAAACTCTGTAAGTGTGAGCGGTGGAAGCGCGCAGCATCT 2800  
QY 821 SerSerAspProProAlaValHisProProLeuProProProLeuArgValThrSerThrAsn 840  
DB 2801 TCGTCAGATCCGCGAGCTGTCCATCCAGCTGCGCCCTCTTCGCGTGACATCTTACCAAT 2860  
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860  
DB 2861 CCGCTGACCCCGCAGCGCCCGCCCGCTTCCCAAGACCGCCAGCGTAAATGGAACCTTG 2920  
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880  
DB 2921 AGCCAGCGGAGAGAGCTGCGCCCGCTGGGATCTCACAGATCAGGCCCCCACCTCTGCC 2980  
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGly 895  
DB 2981 CCACAGCGCCCGCAGCGCCCTCCCGCAGAAAGACCTTCGCGCGGGG 3025

## RESULT 13

US-09-949-003C-925  
; Sequence 925, Application US/09949003C  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC  
; DISORDERS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL000791  
; CURRENT APPLICATION NUMBER: US/09/949,003C  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/231,446  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 74065  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 925  
; LENGTH: 3025  
; TYPE: DNA

i ORGANISM: Homo sapiens

US-09-949-003C-925

## Alignment Scores:

Pred. No.:	5,44e-192	Length:	3025
Score:	4664.00	Matches:	893
Percent Similarity:	99.78%	Conservative:	0
Best Local Similarity:	99.78%	Mismatches:	2
Query Match:	88.45%	Indels:	0
DB:	41	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-949-003C-925 (1-3025)

Qy	1	MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleVal	20
Db	341	ATGCCGACACAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTACAAAGGCG	400
Qy	21	ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	401	CCCACGGCTCCAGCTTCACACCCGCGCAGCGGCGAGTCCCGGACACTGTGGCGGCCATC	460
Qy	41	GluGluAlaLeuAppValAspArgMetValIleuTyrIleValMetLysLysSerValLysAla	60
Db	461	GAGGAGGCTTTGGACGTGCGACCGATGCTTTCTTTACAAAATCAAGAAATCCGTGAAAGCA	520
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu	80
Db	521	ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACCCAGGCTCTGGAG	580
Qy	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys	100
Db	581	AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTCCGTCTCTGAAG	640
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	641	TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAAATATGAAC	700
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	701	AACATAATCTCTCTCCCTTTGGACAGTTTGTCTGAAGGGGGACCTGAAAGGAGTGAAGGG	760
Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle	160
Db	761	GATCTGAAAAGGCTTTTGATAAAGCTTGGAGGACTATGAACAAAATAAACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLysIleHisGlyMetIleArgThrGluIleSerGly	180
Db	821	GAAGAAGGAGAAAAAGAAACACGCCAAGCTTCCATGGGATGATTCGGACTGAAATTAAGCGGA	880
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	881	GCGGAAATTCGCGAAGAGATGAAAGAGAGGAGCGCTCTTCCAGCTACAGATGTGGCAG	940
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	941	TATCTGCTGAAGTCAACGAATCAAGATTAAGAGGGAGTAGATTTACTTCAGAACTCG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAAAATCTTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAGCGCTGGAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAAACTCTTCATTGAAACGCTGTCTACGGATCTTCACAGATCAAAACAGGCCCCAGGAT	1120
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1121	GAAGAAAGAGGACGCTTGATACAGCTTCAGATATTTTGAATCCGATTCGAGGTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGGAGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGGA	1240

Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AACAAGGAAATATGGACCGGAGCGGAAACGGAGCCTCTACAAGAAAGAGTGACGGGATCCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAGGAAATGTTTCAAGTAAAAATGGTTTCTTGACCATATCCATGGT	1360
Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGGCTCTCTCAAGAGCTCAACCTGTAACCTGCCAGGTGAAGACCAACCT	1420
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAAAGAAAGTCTTTTACCTTATTTTACATGACAGAACTTACCACTTTCAAGCTGAA	1480
Qy	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
Db	1481	GATCAACAGGAATGTCAAAATATGGATGCTGTGTCTGCAAAATAGCAAAAGAAAGCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnIleValGlnGluLeuThr	420
Db	1541	AACAATGCATTTAAGGGGATGACAAATCTGAGAAATAAATCAATCGTCCAAAGAACTGACA	1600
Qy	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1661	CGCGCAGATCTACATGGCTTTCCCAACCTGGGCATCTCTGACCTGCATCGAGTGTTC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1721	GGAAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGAGTGCAGTCCCTGACCTTAGATGTA	1780
Qy	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1781	CTGGGAACATCTGAGCTGTCTGCCCAAGATATTTGGGAATGCAGGCTTTAATGAGATC	1840
Qy	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
Db	1841	ATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGCGAGCATGAAT	1900
Qy	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis	540
Db	1901	GCAAGAAGGACTACATCACAGCCCAAGTACATCAGAGGAGATACCGAAGGAAGAGCAC	1960
Qy	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560
Db	1961	GCGGATTAACGGCGGAGAGCTTCAGATCTTTGCGAGGCCGTCNAACGAGAGATATTTTT	2020
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	2021	GGATTTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAAAC	2080
Qy	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
Db	2081	GGACATGAGCGGATGAACCGGCCCTCACCTTCAGTCAGATCCGTGGATCGAACCTCT	2140
Qy	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620
Db	2141	CTTCACATTTAGACTTTTATGTTAGTTCAGAACAGTGGGAACCTTGGATTAACACAGAGGAA	2200
Qy	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
Db	2201	GGCAGCAGCCCTGCACTACTGCTGCTGACCCAGCAATGCGAGTGCCTCAAGTGTCTC	2260
Qy	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
Db	2261	CTGGGGGGAAGGCTCCATCGATGATGAAACAGAGTCCAGAGAGACTCCCTCTGACAT	2320
Qy	661	AlaLysArgLeuLysHisGluHisGluCysGluLeuLeuThrGlnAlaLeuSerGlyArg	680





Qy 350 AsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIle 369  
Db 1280 AACCTGCTAACTGCCAGGTGAGACCACCTCTGAGGAGAGAGAGTGTCTTACCTTATT 1339  
Qy 370 SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMet 389  
Db 1340 TCACATGACAGAACTTACCACCTTCAAGCTGAAGATGAACAGGAATGCAAAATGGATG 1399  
Qy 390 SerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPheLysGlyAspAsn 409  
Db 1400 TCTGTGCTGCAAAATAGCAAGAGAGCTTAAACAAATGCAATTTAAAGGGGATGACAA 1459  
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429  
Db 1460 ACTGAGAAATAACATCGTCAAGACTGCAAGAGGAGATCATCTCAGAAAGTGACAGG 1519  
Qy 430 MetThrGlyAsnAspValCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449  
Db 1520 ATGACGGGCAATGAGTCTGTCTGTGACTGTGGGGGCCAGATCCTACATGCTTCCACC 1579  
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469  
Db 1580 AACCTGGGCATCCTGACCTGCATCGAGTGTTCGGGAATCCACCGAGAGCTGGGGTTCAT 1639  
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489  
Db 1640 TATTCAGGATGACGTCCCTGACCTTAGATTAGTCTGGGAACATCTGAGCTGCTGCTGCC 1699  
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509  
Db 1700 AAGAATAATTGGGAATGACAGCTTTAATCAGATCATGGAATGTGCTTACCAGCTGAGGAC 1759  
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529  
Db 1760 TCAGTCAAAACCAACCCAGGAGCAGCATGATGAATGCAAGAAAGGACTACATCACAGCCAA 1819  
Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549  
Db 1820 TACATCGAGAGAGATAGCAAGGAGAGACGCGGGATTAACGCGCGAGAGCTTCACAGT 1879  
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaLysAlaAspGly 569  
Db 1880 CTTTTCGAGGCGCTCAAAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGCT 1939  
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589  
Db 1940 GTGGATCTTACGGAATAATCCACTGGCCACGAGACATGAGCCGGATGAAACGGCCCTC 1999  
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609  
Db 2000 CACCTTGAGTCAGATCCGTGGATCGAACCTCTCTTCATTTGATGATCTTTTAGTTGAG 2059  
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629  
Db 2060 AACAGTGGGAACCTGGATAAACAGACAGGAAAGGAGCAGCACAGCCCTGCACCTACTGCTGC 2119  
Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIle 649  
Db 2120 CTGACCGCAATGCGGAGTGCCTCAAGTTGCTCTCGCGGGGAGAGGCTTCCATCGAGATA 2179  
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669  
Db 2180 GCAACAGAGTCAGGAGAGACTCCGCTGGACATTTGCAAGCGCTCAAGCAGCAGCAGCTGT 2239  
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689  
Db 2240 GAGGAGCTGCTGACCAAGCTTATCTGGAAGATTTAATCTCAGCTTCAAGTTCAATAT 2299  
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709  
Db 2300 GAATGGCGACTACTCCAGCAAGACCTGGATGAAGTGTATGACGACATGATGAGAAATG 2359  
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729

Db 2360 CAGCCAGTCCCAACCGCGGAGAGACCGGCCCATCAGCTTCTACCAGTGGCTCCCAAC 2419  
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749  
Db 2420 CAGCTTCAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAAACCTTTGCCAAGAGAAG 2479  
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly 769  
Db 2480 CAGAGGGCTTTCATGCCCAGCATCTTGAGAAATGAGACTTACGGAGGCCCTCTGAGTGGC 2539  
Qy 770 SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProPro 789  
Db 2540 AGCCCACTCCGCGCCAGCTGAGCCGCCAGCACCCAGCAGCGCCCGCTTCTCTCCA 2599  
Qy 790 ArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSer 809  
Db 2600 CGGAATGTTGCCAAA----- 2614  
Qy 810 ValSerValAspGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisPro 829  
Db 2614 ----- 2614  
Qy 830 ProLeuProProLeuArgValThrSerThrAsnProLeuThrProThrProProPro 849  
Db 2615 -----GATCCCTGACCCCGCCAGCGCCCGCCCGCT 2644  
Qy 850 ValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProPro 869  
Db 2645 GTTGCCAAAGCGCCAGCGTAATGGAAGCTTTAGCCAGCGAGCAAGCCCTGCCCGCT 2704  
Qy 870 GlyIleSerGlnIleArgProProProProLeuProGlnProProSerArgLeuProGln 889  
Db 2705 GGGATCTCACAGATCAGGCCCCCACCCTCTGCCCCACAGCCCGCCAGCGCTTCCCGAG 2764  
Qy 890 LysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909  
Db 2765 AAGAAGCTGCGCGGG----- 2782  
Qy 910 GlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeu 929  
Db 2783 -----TTGGATCTCTGCAACGGAAGCTCTGGTCTCTGCTCAATGCTTATGTCCTG 2836  
Qy 930 GlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVal 949  
Db 2837 CAGCCCCCTGACCCATGCTTAGAAGTCGCGAGCGAACCAAGTTGAGCCCTAAGCGGCTG 2896  
Qy 950 LysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAsp 969  
Db 2897 AAGCGCTCTATAACTGTGTGCTGACAAACCCGATGAGCTCACCTTCTCCGAGGGGAT 2956  
Qy 970 ValIleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAsp 989  
Db 2957 GTGATCATCTGTCGCGGGAGGAGCCAGAGAGTGGTGGATGGCCACATGATGGAGAT 3016  
Qy 990 ProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006  
Db 3017 CCGTGCAGAAAGCGGCATTCGCGTGCATTTGTGCACCTTTATCGCTGAC 3067

## RESULT 15

US-60-452-680-5750  
; Sequence 5750, Application US/60452680  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001450  
; CURRENT APPLICATION NUMBER: US/60/452,680  
; CURRENT FILING DATE: 2003-03-07  
; NUMBER OF SEQ ID NOS: 116213  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5750



```

; LENGTH: 5427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-5750

Alignment Scores:
Pred. No.:      2,298-187      Length:      5427
Score:          4562.50      Matches:      892
Percent Similarity: 87.91%      Conservative: 2
Best Local Similarity: 87.71%      Mismatches: 4
Query Match:      86.53%      Indels:      119
DB:              115          Gaps:        5

US-09-914-042-1 (1-1006) x US-60-452-680-5750 (1-5427)

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
DB 341 ATGCCGGACCATCTCCGCTCGAATTCGTGGCGGAGACCCATGAGGACTACAGGCG 400
QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 40
DB 401 CCCAGGCGCTCCAGCTTCACCAACCGCACGGCGCAGTGCAGAACACACTGTGGCGCCATC 460
QY 41 GluGlu-----AlaLeuAspValAspArgMet 49
DB 461 GAGGAGGTGAGCGCGCTCGCGCGGCTCGCGCGCAGGCTTTGGAYGTGGACCGGATG 520
QY 50 ValLeuTyrLysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisVal 69
DB 521 GTTCTTTACAAATGAGAAATCCGTGAAGCAATCAACAGCTCTGGGCTGGCTCACGTG 580
QY 70 GluAsnGluGluGlnTyrThrGlnAlaLeuGluLysPheGlyGlyAsnCysValCysArg 89
DB 581 GAAATGAGAGCAGCTACACCCAGGCTCTGGAGAAAGTTTGGCGGCAACTGTGTATGAGA 640
QY 90 AspAspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
DB 641 GATGACCCAGATTTAGGAAGTTCGTCTCAGAGTTCTCAGTGTTTACAAAGGAGTTGACA 700
QY 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSer 129
DB 701 GCACITTTCAAAACCTG-----718
QY 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysPheAspLysAla 149
DB 719 -----GATCTGAAAAGSCTTTTGATAAGCT 745
QY 150 TrpLysAspTyrGluThrLysIleThrLysIleGluLysGluLysGluLysGluHisAlaLys 169
DB 746 TGGAAAGGACTATGAACCAAAATAACCAAGATAGAAAAGGAGAAAAGGAAACACGCCAAG 805
QY 170 LeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLys 189
DB 806 CTCCTATGGATGATTCGGACTGAAATAGCGGAGCGGAAATTCGCCAGAGAGATGAAAAG 865
QY 190 GluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLys 209
DB 866 GAGAGCGCTCTCTCCAGCTACAGATGTCGAGTATCTGCTGAAGTCAACGAAATCAAG 925
QY 210 IleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsn 229
DB 926 ATTAAGAGGGAGTAGATTACTTCAGAAATCTGATCAAAATACCTTTTCATGCCCAATGCAAT 985
QY 230 PhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSer 249
DB 986 TTTTTCAGGATGGACTCAAGCCGTGGAAGCCCTCAAAACCTTCCTTCAAGACCGTGTCT 1045
QY 250 ThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeuIleGlnLeu 269
DB 1046 ACGGATCTTCACCGATCAAAACAGGCCCGGATGAGAAAGAGGAGGAGTGTGATACAGCTT 1105
QY 270 ArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGln 289

```

```

DB 1106 CGAGATATTTGAAATCCGCATTCGAGGTTGAAACAGAAAGAGGACTCCCAAAATTCGTGAC 1165
QY 290 SerThrAlaTyrSerLeuHisGlnProGlnGlnLysGluHisGlyThrGluArgAsn 309
DB 1166 AGCAGAGCTTATAGCTTACATCAGCTCAGGGAACAAGGAACATGGGACCGAGCGGAAC 1225
QY 310 GlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSer 329
DB 1226 GGCAGGCTCTCAAGAAGAGTGCAGGGTAC-----1255
QY 330 ValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu 349
DB 1256 -----GCTAACCGGCCCTCTCTCAAAAGCTC 1279
QY 350 AsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIle 369
DB 1280 AACCTGCTAACTGCCAGGTGAAGCAACCACTCAGGAGAGAGAGTGTCTTGACCTYATT 1339
QY 370 SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMet 389
DB 1340 TCACATGACAGAACTTACCACTTTCAAGCTGAAGATGAACAGGAATGTCAAATATGGATG 1399
QY 390 SerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsn 409
DB 1400 TCTGTGCTGCAAAATAGCAAAAGAAAGACTTTAAACAATGCAATTTAAGGGGGATGACAAAT 1459
QY 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
DB 1460 ACTGGAGAAATACATCGTCCAGAACTGACAAAGAGAGATCATCTCAGAGTGCAGAGG 1519
QY 430 MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
DB 1520 ATGACGGCAATGAYGTCTGTGCTGCTGCGGCGCCAGATCTCTACATGGCTTTCCACC 1579
QY 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
DB 1580 AACCTGGGCACTCCGACCTGCATCGAGTGTTCGGGAATCCACCGAGAGCTGGGGGTTCAT 1639
QY 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
DB 1640 TATTCAGGATGCAGTCCCTGACCTTAGATGTACTGGGNAACATCTGAGCTGCTGCTCGCC 1699
QY 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
DB 1700 AAGAATATTTGGAAATGCAAGCTTTAATGAGATCATGGAATGTGCTCACCAGCTGAGGAC 1759
QY 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
DB 1760 TCAGTCAAAACCCACCCAGGCGAGCATGAATGCAAGAAAGGACTATCATCACGCCAAG 1819
QY 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
DB 1820 TAYATCGAGAGAGATAGCAGAGGAGAGAGGACGCGGATTAACGCGCGGAGGTTTACAGT 1879
QY 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
DB 1880 CTTTTCGAGGCGGTCAAAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGT 1939
QY 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
DB 1940 GTGGATCTTACGGAATAAATCCCTGCGCAACGAGACATGAGCCGGATGAAACCGCCCTC 1999
QY 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
DB 2000 CACCTTGGCGTTCAGATCCGTGGATCGAACTCTCTTTCATATTGTAGACTTTTATTGTTAG 2059
QY 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
DB 2060 AACAGTGGAACTTGGATAAACAAGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2119
QY 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 649
DB 2120 CTGACCGCAATTCGCGAGTGCCTCAAGTTGCTCTCGCGGGGAGGAGGAGGAGGAGGAGGAGGAG 2179

```

Search completed: August 4, 2005, 16:31:38  
Job time : 8888 secs

```
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 2180 GCAAAYAGAGTCAGGAGAGACTCCGCTGACATTTGCCAAGCGCTCAAGCAGCAGCAGCTGT 2239
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2240 GAGGAGCTGCTGACCAAGCCTTATCTGGAAGATTTAATTTCTCAGCTTTCACGTTGCAATAT 2299
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2300 GAATGGCGACTACTCCAGCAAGACTGGATGAAGTATGACGACATGATGAGAAATG 2359
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2360 CAGCCAGTCCCAACCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAC 2419
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2420 CAGCTTCAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTTGCCAAGGASAAG 2479
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly 769
Db 2480 CAGAGGGCTTTTCATGCCAGCATCTTGAGAATGAGACTTACGGAGCCCTCCTGAGTGGC 2539
Qy 770 SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProLeuProPro 789
Db 2540 AGCCCACTCCCGCCAGCTGACGCCCCAGCACACAGCGCCCCCGCTTCTCTCCA 2599
Qy 790 ArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSer 809
Db 2600 CGGAATGTTGCCAAA----- 2614
Qy 810 ValSerValAspGlyGlySerArgGlnArgSerSerSerAspProAlaValHisPro 829
Db 2614 ----- 2614
Qy 830 ProLeuProProLeuArgValThrSerThrAsnProLeuThrProThrProProPro 849
Db 2615 -----GATCCCTGACCCCCCAGCGCCGCCACCC 2644
Qy 850 ValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProPro 869
Db 2645 GTTGCAAGAGCGCCAGCGTAATGAGACCTTGAGCCAGCGGAGCAAGCCTGCCCGCCT 2704
Qy 870 GlyIleSerGlnIleArgProProProLeuProGlnProProSerArgLeuProGln 889
Db 2705 GGGATCTCACAGATCAGGCCCCCACCCTCTGCCCCACAGCGCCCGCCGCTCCCGCAG 2764
Qy 890 LysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909
Db 2765 AAGAAGCTGCGCCGGG----- 2782
Qy 910 GlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeu 929
Db 2783 -----TTGGATCTCTGCAACGGAAGCTCTGGGTCTCTGTCCAATGCTATGTCCTG 2836
Qy 930 GlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVal 949
Db 2837 CAGCCCTCTGACCCATGCTAGGAAGTCGCAAGCAACCAAGTTGAAGCCTAAGCGGGTG 2896
Qy 950 LysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAsp 969
Db 2897 AAAGCGCTCTATATCTGTGTGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGAT 2956
Qy 970 ValIleIleValAspGlyGluGluAspGlnGluTyrTrpIleGlyHisIleAspGlyAsp 989
Db 2957 GTGATCATCGTGACGGGAGGAGGAGCCAGAGATGGTGGATTGGCCACATTGATGAGAT 3016
Qy 990 ProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3017 CCTGGTCGCAAGGGCGCATTCGCGGTGTCATTGTGTGCATTTTATYGTCTGAC 3067
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:58:17 ; Search time 2694 Seconds  
(without alignments)  
2420.646 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAPFVSFVHFAD 1006

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7297361 seqs, 3241162794 residues

Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool\_h/US09914042/runat\_04082005\_090635\_2186/app.query.fasta\_1.1159  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09914042 @CGN 1.1 658 @runat\_04082005\_090635\_2186  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGEQUERY -NEG\_SCORE=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5253	99.6	5711	21	US-10-956-157-1144	Sequence 1144, Ap
2	3006	57.0	3974	18	US-10-210-281-73	Sequence 73, Appl
3	2371.5	45.0	5582	17	US-10-062-674-1606	Sequence 1606, Ap
4	2089.5	39.6	3812	17	US-10-104-047-981	Sequence 981, Appl
5	2089.5	39.6	4050	21	US-10-490-605-1	Sequence 1, Appl
6	1801.5	34.2	3436	11	US-09-764-875-358	Sequence 358, Appl
7	1759.5	33.4	3413	11	US-09-764-875-358	Sequence 37, Appl
8	1307.5	24.8	3123	17	US-10-104-047-314	Sequence 314, Appl
9	695	13.2	563	9	US-09-815-343-1154	Sequence 1154, Ap
10	695	13.2	563	18	US-10-097-105-1154	Sequence 1154, Ap
11	571.5	10.8	6990	15	US-10-037-270-155	Sequence 155, Appl
12	571.5	10.8	6990	17	US-10-117-722-155	Sequence 155, Appl
13	570.5	10.8	2484	17	US-10-080-334-79	Sequence 79, Appl
14	561	10.6	2737	15	US-10-176-306-10	Sequence 10, Appl
15	558	10.6	5421	20	US-10-723-860-8033	Sequence 8033, Ap
16	532.5	10.1	2505	15	US-10-176-306-12	Sequence 12, Appl
17	532.5	10.1	2505	18	US-10-467-434-22	Sequence 22, Appl
18	519.5	9.9	2810	20	US-10-739-930-640	Sequence 640, Appl
19	509	9.7	6482	21	US-10-956-157-4819	Sequence 4819, Ap
20	492.5	9.3	2039	9	US-09-822-830A-256	Sequence 256, Appl
21	481	9.1	469	18	US-10-276-774-288	Sequence 288, Appl
22	479.5	9.1	399	17	US-10-442-535A-21815	Sequence 21815, A
23	479.5	9.1	399	18	US-10-085-783A-21815	Sequence 21815, A
24	453.5	8.6	2984	20	US-10-739-930-745	Sequence 745, Appl
25	441.5	8.4	2773	19	US-10-437-963-34050	Sequence 34050, A
26	429	8.1	2041	9	US-09-828-303-11	Sequence 11, Appl
27	429	8.1	2041	19	US-10-716-089-11	Sequence 11, Appl
28	428.5	8.1	442	10	US-09-918-995-27010	Sequence 27010, A
29	424	8.0	326	9	US-09-815-343-1151	Sequence 1151, Ap
30	424	8.0	326	18	US-10-097-105-1151	Sequence 1151, Ap
31	416	7.9	2331	21	US-10-921-707-19	Sequence 19, Appl
32	403	7.6	409	18	US-10-276-774-404	Sequence 404, Appl
33	396.5	7.5	2708	19	US-10-437-963-78764	Sequence 78764, A
34	381.5	7.4	1950	19	US-10-437-963-79456	Sequence 79456, A
35	388.5	7.4	2982	10	US-09-814-353-19609	Sequence 19609, A
36	387.5	7.3	451	17	US-10-062-674-482	Sequence 482, Appl
37	357.5	6.8	1764	17	US-10-364-237-1383	Sequence 1383, Ap
38	340.5	6.5	3143	17	US-10-334-143-108	Sequence 108, Appl
39	336.5	6.4	3950	17	US-10-172-118-1530	Sequence 1530, Ap
40	336.5	6.4	3950	18	US-10-342-887-1530	Sequence 1530, Ap
41	333.5	6.3	3314	18	US-10-424-599-17418	Sequence 17418, A
42	330.5	6.3	2730	17	US-10-094-749-373	Sequence 373, Appl
43	329	6.2	1803	20	US-10-425-115-48371	Sequence 48371, A
44	328.5	6.2	4768	20	US-10-719-993-411	Sequence 411, Appl
45	328.5	6.2	4792	20	US-10-719-993-410	Sequence 410, Appl

# ALIGNMENTS

RESULT 1

US-10-956-157-1144  
; Sequence 1144, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956.157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1144  
; LENGTH: 5711

```
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-956-157-1144

Alignment Scores:
Pred. No.: 0 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 21 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
Db 341 ATGCCGGACCAAGATCTCCGTGTGCGAATTCGTGGCCGAGACCCATGAGGACTCAAGGCG 400

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGCCTCCAGCTTCACCACCCGCGACGGCGACGTCCCGAACACTGTGGCGCCATC 460

Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60
Db 461 GAGGAGGCTTTGGACGTGACCGGATGCTTCTTACAAATGAAGAAATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCGTGGCTCAGCTGCGAATGAAGAGCAGTACACCCAGGCTCTGGAG 580

Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheIleLys 100
Db 581 AAGTTTGGCGGCAACTGTGTATGACAGATGACCCAGATTTAGGAAGTGCCTCTCGAAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTTGATTCAGAAATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
Db 701 AACATAATCTCTCCCTTTGACAGTGTGCTGAAGGGGGACCTGAAGAGGAGTGAAAGGG 760

Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrIlystle 160
Db 761 GATCTGAAAAAGCCTTTTGATAAAGCTTGGAGGACTATGAACAAAAAATAACCAAGATA 820

Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAAAAAGCGGA 880

Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGAAAGAGAGAGGCGCTTCTTCCAGCTACAGATGTGCGAG 940

Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGCTGAAGGTCACGAATCAAGATTAAGATTAAGAGGAGTAGATTTTACTTCAGAAATCTG 1000

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAAGCCGTGGAAAGC 1060

Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTCTCCATTTGAAACGCTGTCTACGGATCTTTCACAGATCAAACAGAGCCCGAGGAT 1120

Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGAGCTTGATACAGCTTCGAGATATTTTGAATCCGCATTGCAGGTTGAA 1180

Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAATTCGTTCAGAGCAGCAGCTTATAGCTTATACAGCCTCAGGGA 1240
```

```
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrIlyLysSerAspGlyIleArg 320
Db 1241 AACAAAGAAACATGGGACCGGACCGGAAACGGCAGCTCTACAAAGAAAGAGTGACGGGATCCGA 1300

Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAGTGTGGCAGAAAGAAAGAAATGTTCAATTAAAATGGTTTCTTGACCATATCCATGGT 1360

Qy 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAGCTCAACTGCTTAACCTGCCAGGTGAAGACCAACCT 1420

Qy 361 GluLylLysCysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAGAAGTGTCTTGACCTTATTTTCAATACAGAACTTACCACTTTTCAAGCTGAA 1480

Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATGAACAGGAAATGTCAAATATGGATGCTGTGTCGCAAAATAGCAAAAGAAAGCTTTA 1540

Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACAAATGCAATTAAGGGGATGACAACTACTGGAGAAAAATAACATCGTCCAAGAACTGACA 1600

Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATGACGCTCTGTGTGCTGTGGG 1660

Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCCAGATCTCTACATGGCTTTCACCAACCTGGGCATCTGTGACCTGCATCGAGTGTCC 1720

Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuLeuAspVal 480
Db 1721 GGAATCCACCAGAGCTGGGGGTTTCAATTTATCCAGGATGCAGTCCCTGACCTTAGATGTA 1780

Qy 481 LeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGAAACATCTGAGCTGCTGCTCGCAAGAAATATGGGAATGCAAGGCTTTAATAGAGATC 1840

Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTGCCTTACCAGCTGAGGACTCAGTCAAAACCAACCCAGCAGGACATGAAT 1900

Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGAACAC 1960

Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 GCGGATAACCGCGGAGGCTTTCACAGTCTTTTGGAGGCGCTCAAAACAGAGAGATATTTTT 2020

Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATTTGCTCCAGCTTATGCTGATGTGTGATCTTACGGGAAAAAATCCCACTGGCCCAAC 2080

Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCCGATGAAACGGCCCTCACCTTCAGTCAGTCAGATCCGTGGATCGAACCTCT 2140

Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2141 CTTTCAATGTTAGACTTTTTTAGTTTCAGAACAGTGGGAACCTGGGATAAACACAGAGGAAA 2200

Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLysLeuLeuLeu 640
Db 2201 GGCAGCACAGCCCTGCACACTACTGCTGCTGACCCGACCAATGCCGAGTGCCTCAAGTTGCTC 2260

Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGGAAGGCTCTCATCGATAGTACAAACGAGTCCAGGAGAGACTCCCGCTCGACATT 2320
```

```
QY 661 AlaLysArgLeuLysHisCysGluGluLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAAGCAGACACTGTGAGGAGCTGCTGACCCAAAGCCTTATCTGGAAGA 2380
QY 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAATCTCAGTTTCAGTTGAAATATGATGGCACTACTCCAGAGACCTTGGATGAA 2440
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db 2441 AGTGATGACGACATGGATGAGAAATTCAGCCCAAGTCCCAACCGCGGGAAGACCGGCC 2500
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTTACCACTGGCTGCCAACCAAGCTTCAGTCTTAACGCTGTATCTTTGGCCAGA 2560
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCNAACCTTGCCNAGGAGAGCAGAGGCTTTTCATGCCCAGCATCTTGCAGAAAT 2620
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGGAGCCCTCTGAGTGGCAGCCCACTCCCGCCCGCCGCTGCAGCCCGCCAGC 2680
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCGCCGCTTCTCCACGGAATGTTGGCAAGTTCAGACAGCCTCTCT 2740
QY 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTTGTGAAGACAAACTCTGTAAAGTGGAGCGGTGGAGCGCGGAGCGGATCT 2800
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCTCTTCGGGTGACATCTACCAAT 2860
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCCAGCCCGCCCGCTTGCAGAGAGCCCGCCAGCGTAAATGGAAGCCTTG 2920
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGGAGCAAGCCCTGCCCGCTGGGATCTCACAGATCAGGCGCCCGCCCTCTGCC 2980
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCCAGCGCTCCCGCAGAGAGCTTGGCGGGGGCTGCACAGTCCAC 3040
QY 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCTG 3100
QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGCTCTCGAGCGCCCTGCACCCATCCCTAGGAAGTCGCGAG 3160
QY 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAACCTTAAGCGGGTGAAGGGCTCTATTAAGTGTGGCTGACAAACCC 3220
QY 961 AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCGTGGAGCGGGAGGAGGACGAGGAG 3280
QY 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTCGATGGCCACATTGATGAGATCCTGTGTCGCAAGGCGCATCTCCCGGTGTCATTT 3340
QY 1001 ValHisPheIleAlaAsp 1006
Db 3341 GTGCACCTTATTCGTGAC 3358
```

RESULT 2

```
US-10-210-281-73
; Sequence 73, Application US/10210281
; Publication No. US20040030096A1
; GENERAL INFORMATION:
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Zhong, Mei
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles E.
; APPLICANT: Ji, Weizhen
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Sciore, Paul
; APPLICANT: Stone, David J.
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Casman, Stacie
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc L.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: 21402-416D
; CURRENT APPLICATION NUMBER: US/10/210,281
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: 60/309,501
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/310,291
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/361,775
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/361,832
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,203
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,201
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,702
; PRIOR FILING DATE: 2001-08-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 73
; LENGTH: 3974
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)..(3656)
US-10-210-281-73
Alignment Scores:
Pred. No.: 1.86e-217 Length: 3974
Score: 3006.00 Matches: 626
Percent Similarity: 66.58% Conservative: 127
Best Local Similarity: 55.35% Mismatches: 234
Query Match: 57.01% Indels: 144
DB: 18 Gaps: 22
US-09-914-042-1 (1-1006) x US-10-210-281-73 (1-3974)
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 321 ATGCGGACCAAGATCTCTGTCTCGAGTTCTATCGCCGAGACCCGCGAGACTTACAACTCG 380
```

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
Db 381 CCCACCACGTCACGCTCCACACGCGGCTGCACAACTGCAGGAAACACCGTCACGCTGCTG 440  
Qy 41 GluGluAlaLeuAapValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
Db 441 GAGAGGCTCTAGGCCAAGATAGAACAGCCCTTCAGAAAGTGAAGAGTCTGTAAAGCA 500  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
Db 501 ATATATAATTCTGGTCAAGATCATGTACAAATGAAGAAACTATGCAACAAGTCTTGAT 560  
Qy 81 LysPheGlyLysAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
Db 561 AAGTTGGAGGTAATTTTTTAAGTCGACAGAACCCCGACCTTGGCAGCGCGTTGTCAAG 620  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120  
Db 621 TTTTCTACTCTTACAAAGAACTGTCCACACTGCTGTAATAATCTGCTCCAGGTTTGAGC 680  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 681 CACAATGTGATCTTCACCTTGGATTTCTTTTAAAGGAGAGACCTAAAGGGAGTCAAAGGA 740  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 741 GATCTCAAGAAAGCCATTTGACAAAGCCCTGGAAGATTTATGACAAAGTTTACAAAAATT 800  
Qy 161 GluLysGluLysLysGluHisAlaLysLysGlyMetIleArgThrGluLysSerGly 180  
Db 801 GAGAAAGAGAAAGAGACAGACGCAAAACACATGGGATGNTCCGCACAGAGATACAGGA 860  
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
Db 861 GCTGAGATTGCGGAAGAAATGAGAAGGAAAGGCGCTCTTTCAGCTCCAAATGTGTGAA 920  
Qy 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 921 TATCTCATTAAGTTAATGAAATCAAGACCAACAAAAGGGGTGGATCTCTCGAGAATCTT 980  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 981 ATAAAGTATTACCATGCAGTGCATTTCTTCAAGATGGCTTGAAACACAGCTGATGAAG 1040  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 1041 TTGAACACAGTACATTGAAATACTGCTGCTGATTTATATATATAATAATAAAACAGACCCAGGAT 1100  
Qy 261 GluGluArgArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 1101 GAAGAAAGAAACACGCTAACTCCAGACTTAATAAATCTCTCTTCAACTGGAT 1160  
Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
Db 1161 CAGAAAGAAATCTAGAGAGATTCTCAGAGCGGCGAAGCA---GGATACAGATGCTATCAG 1217  
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
Db 1218 CTCACGGGCAATAGGAATATAGGCAGTGAAAGAGAGGGGTACCTGCTAAAGAAAGTGCAC 1277  
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
Db 1278 GGGATCCGGAAGATATGCGAGAGGAGGAAGTGTTCAGTCAAGAAATGGGATTTCTGACCATC 1337  
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
Db 1338 TCACATGCCATCTTAACAGGCAACACGCAAGTTGAACCTTCTCACCTGCCAAGTAAAA 1397  
Qy 358 ThrAsnProGlnLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377  
Db 1398 CCTAATGCCGGAAGACAAAAAATCTTTTGACCTGTATATACATAATAGAACATATACATTT 1457

Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
Db 1458 CAGCAGCAAGATAGCAGGATTTATGTAGCATGGATATCATGATTGACAAATAGCAAGA 1517  
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAspThrGlyGluAsnAsnIleValGln 417  
Db 1518 GAGGCCCTAACCATGGCTTCCGTGGAGAGCAGAGTGGGGAGAGAGACAGCTG---GAA 1574  
Qy 418 GluLeuThrLysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
Db 1575 GACCTGCAAAAGCCATTATTAGGATGTCAGCGGCTCCAGGGATCCAGGGAATGACATTTGCTGC 1634  
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457  
Db 1635 GATTGTGGCTCATCAGAACCCACCTGCTTTCACCAACCTTGGGTATTTTTCACCTGTATA 1694  
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
Db 1695 GAATGTTCTGCATCCATAGGGAATGGGGTTTCATATTCTCGCATTCAGTCTTTGGAA 1754  
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
Db 1755 CTAGACAAATTAGGAACCTTCTGAACCTCTGTGGCCAGAAATGTAGGAAACAATAGTTT 1814  
Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
Db 1815 AATGATATTATGGAAGCAAAATTTTACCAGC---CCCTCACCAAAACCCACCCCTTCAAGT 1871  
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArg 537  
Db 1872 GATATGACTGTACGAAAGAAATATATCACTGCAAAGTATGTAGATCATAGGTTTCAAGG 1931  
Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
Db 1932 AAGACCTGTTCAACTTCATCAGCTAAACTAAATGAATTCGTTGAGGCCATCAAAATCCAGG 1991  
Qy 558 AspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
Db 1992 GATTTACTTGTCACATAATTCAGTCTATGCAAGAGGGGTAGAGCTAAATGGAAACCACTG--- 2048  
Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
Db 2049 CTGGAACCTGGCAGGAGCTTGGGAGACAGCCCTTCACCTTGGCGTCCGAATGCAGAT 2108  
Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
Db 2109 CAGACATCTCTCCATTTGGTTGACTTCTCTGTACAAAACTGTGGGAACTCTGGATAAGCAG 2168  
Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeu 637  
Db 2169 ACGCCCTGGGAACACAGATTCTACACTACTGTATGTATGTACAGTAAACCTGAGTGTGTTG 2228  
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657  
Db 2229 AAGCTTTTGCTCAGGACGACCCCACTGTGGATATAGTTAAACAGGCTGGAGAACTGCC 2288  
Qy 658 LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677  
Db 2289 CTAGACATAGCAAGAGACTAAAAAGTACCAGTGTGAAGATCTGCTTCCAGGCTAAA 2348  
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697  
Db 2349 TCTGGAAGTTTCAATCCACAGCTCCAGTAGAATATGATGGAATCTTCGACAGGAGGAG 2408  
Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlnAsnArgGlu 717  
Db 2409 ATAGATGAGCGCATGATGATCTCGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCA 2468  
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737  
Db 2469 CCCAGACCTCAGAGCTTCTGCACTCTCTCCAGCATCTCCCCCCAGGAC-----AAG 2519  
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757

```
Db 2520 CTGGCACTGCCAGGATTGAGCACTCAAGGGGCAAAACAGCGGCTC----- 2564
Qy 758 LeuGlnAenGluThrTyrGlyAlaLeuLeuSerGly-----SerProProAla 774
Db 2565 -----TCCTATGGAGCTTCCACCAACAGATCTTCGTTTCCACAGACACAC 2612
Qy 775 GlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794
Db 2613 TCGCCACATCACCA---ACCACGAGGCTCCCTCTGCCCCCTAGGAACGCCGGGAAA 2669
Qy 795 -----ValGlnThrAlaSerSerAlaAsn 802
Db 2670 GGTCAACTGGCCCACTTCAACACTCCTCTAAGCACCCAGACCTAGTGGCAGCTCC 2729
Qy 803 ThrLeuTrpLysThrAsnSerValSerValAepGlyGlySerArgGlnArgSerSer 822
Db 2730 ACCCTATCCNAGAAGAGCGCTCTCCCTCCACCCAGCCGGA---CACAGAGAACCTATCC 2786
Qy 823 AspProAlaValHisProProLeuProProLeuArg----- 835
Db 2787 GACCTCCAGCCCACTACCTCATGGGCCCCCAAAACAAAGGCGCAGTTCTCTGGGGTAAC 2846
Qy 835 ----- 835
Db 2847 GATGGGGTCCATCTCTTCAAGTAAGACTACAAACAAGTTTGAGGGACTATCCAGCAG 2906
Qy 836 -----ValThrSerThrAsnPro---LeuThrProThrProProPro 848
Db 2907 TCGAGCACGAGTTCTGCAAGACTGCCCTTGCCCAAGAGTTCTTCTAAACTACTCAG 2966
Qy 849 ProValAla-----LysThr-----ProSer 855
Db 2967 AAAGTGGCACTAAGGAAAAAGATCATCTCTCCCTAGACAAAGCCACCATCCCGCGGAA 3026
Qy 856 ValMetGluAlaLeuSerGln-----ProSerLysProAlaPro----- 868
Db 3027 ATCTTTCAGAAATCATCAGATGTCAGAGTTGCCCAAAAGCCACCACTCGGAGACCTG 3086
Qy 869 -----ProGlyLysSerGlnLeuArgProProPro 878
Db 3087 CCCCCAAAGCCACAGAACTGGCCCCCAAGCCCCCAATTTGGAGATTTGCCGCTTAAGCA 3146
Qy 879 -----LeuProProGlnProPro---SerArg 886
Db 3147 GGAGAACTGCCCCCAACACACAGCTGGGGAGCTGCCACCAACCAACCACTCTCAGAC 3206
Qy 887 LeuProGlnLys----- 890
Db 3207 TTACCTCCCAAAACACAGATGAAGGACCTGCCCCCAACCAACACAGCTGGGAGACCTGCTA 3266
Qy 891 -----LysProAlaProGlyThrAsp 897
Db 3267 GCATAATCCAGACTGGAGATGCTCTACCCAGGCTCAGCAACCTCTGAGGTACACTG 3326
Qy 898 LysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSer 915
Db 3327 AAGTCACACCACTTGATCTATCCCAATGTGCACTCCAGAGACGCCATCCAAAGCAA 3386
Qy 916 AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMet 935
Db 3387 GCATCTGAAGACTTCAACGACCTACCTCTGTGCA---GAGACGCCGTATCACTG 3443
Qy 936 ProArgLysSerGlnAlaThrLysLysLysProLysArgValLysAlaLeuTyrAsnCys 955
Db 3444 CCGAGAAAATCAATACGGGGAAAAATAAGTGGCGGAGTGAACCATTTATGACTGC 3503
Qy 956 ValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGly 975
Db 3504 CAGGACAGACAGGATGACGAGCTCACATTCTCAGAGGAGGAAGTATTCGTACACGG 3563
Qy 976 GluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAla 995
```

```
Db 3564 GAAGAGGACCAAGGAGTGTGGATTGGCCACATCGAAGGACACACCTGAAAGAGGGGGTTC 3623
Qy 996 PheProValSerPheValHisPheIleAlaAsp 1006
Db 3624 TTTCAGTGTCTTTGTTTCATATCTCTGCTGAC 3656

RESULT 3
US-10-062-674-1606
; Sequence 1606, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: PA-0036-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/625,102
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 1606
; LENGTH: 5582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040005559A1 199975.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ... (5582)
; OTHER INFORMATION: a, t, c, g, or other
US-10-062-674-1606

Alignment Scores:
Pred. No.: 3,4e-169 Length: 5582
Score: 2371.50 Matches: 509
Percent Similarity: 62.40% Conservative: 105
Best Local Similarity: 51.73% Mismatches: 194
Query Match: 44.97% Indels: 176
Dbs: 17 Gaps: 22

US-09-914-042-1 (1-1006) x US-10-062-674-1606 (1-5582)
Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIle 183
Db 2 AAAAGAGAGCAGCAAAACAAACATGGGATGATCCGACAGAGATAACAGGAGCTGAGATT 61
Qy 184 AlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeu 203
Db 62 GCGGAAGAAATGGAGAGAAAGCGCTCTTTTCAGCTCCAAATGTGTGAATATCTCATTT 121
Qy 204 LysValaenGluIleLysLysGlyValaLysGlyValaLysGlyValaLysGlyVala 223
Db 122 AAAGTTAATGAATCAAGACCAAAAGGTTGGATCTGCTGCAGATCTTTATAAAGTAT 181
Qy 224 PheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243
Db 182 TACCATGCACAGTGCATTTCTTTTCAGATGCTTGAACACAGCTGATAAGTTGAACAG 241
Qy 244 SerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArg 263
Db 242 TACATTGAAAAAATCGCTGCTGATTTATATAATAATAAAACAGACCAGGATGAAGAAAG 301
Qy 264 ArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGlu 283
Db 302 AAACAGCTAACTGCACCTCCGAGACTTATAAAATCTCTCTTCAACTCGATCAGAAAGAA 361
Qy 284 -----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 362 TCTAGGAGAGATTCTCAGAGCCGCGCAAGGA---GGATACAGCATGTCATCCAGGGC 418
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArg 320
```

Db 419 AATAAGGAATATGGCAGTGAAGAGGGGTACCTGCTAAAGAAAGTGAAGGGATCCGG 478  
Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 479 AAAGTATGGCAGAGAGAGAGTGTTCAGTCAAGAAATGGGATTTCTAACCATCTCAATGCC 538  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 539 ACATCTACAGGCAACACGACAGTTGAACCTTCTCACCTGCGCAAGTAAACCTTAATGCC 598  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 599 GAAGACAAAATAATCTTTTACCTGATATCATATAATAGAAATATATCACTTTTCAGGCAGAA 658  
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 659 GATGACGAGGATATGTAGCATGGATATCAGTATTGACAAATAGCAAAAGAGAGGCCCTA 718  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 719 ACCATGGCTTCCGTGGAGAGCAGAGTGGGGAGAGAACAGCCTG---GAAGACCTGACA 775  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 776 AAAGCCATTTAGGATGTCCAGGGCTCCAGGGGAATGACATTTGCTGGATTTGGGC 835  
Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGluIleLeuThrCysIleGluCysSer 460  
Db 836 TCATCAGAACCCACCTGGCTTTCAACCAACTTGGGTATTTTACCTGTATAGATGTTCT 895  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 896 GGCATCCATAGGAAATGGGGTTCATATTCTCGCATTCAGTCTTTGGAACCTAGACAAA 955  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 956 TTAGGAACCTTCTGAACCTTCTGCGCCAAAGATGTAGGAACAATAGTATTTTAATGATATT 1015  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1016 ATGGAAGCAAAATTTACCAGC---CCCTCACCAAAACCCACCCCTTCAAGTGATATGACT 1072  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db 1073 GTACGAAAGAAATATATCACTGCAAAAGTATGTAGATCATAGTGTTCAGGAAGACCTGT 1132  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1133 TCAACTTCATCAGCTAAACTAAATGAATGCTTGAGGCCATCAAAATCCAGGGATTTACTT 1192  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1193 GCACCTAATTCAGTCTATGCAAGAGGGGTAGAGCTTAATGGAACCACTG---CTGGAACCT 1249  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 1250 GGGCAGGAGCTTGGGAGACGCCCTTCACTTGCCTGCGGACTCGCAGATCAGACATCT 1309  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 1310 CTCATTTGGTTGACTTCTCTGTACAAAACCTGGGAACTTGGATAAGCAGAGCGCCCTG 1369  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 1370 GGAACACAGATTTCTACACTACTGTATGTATGTACAGTAAACCTTGAGTGTTTGAAGCTTTG 1429  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 1430 CTCAGGACCAAGCCCACTGTGGATATAGTTAACAGGCTGGAGAACTGCCCTTAGACATA 1489  
Qy 661 AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 1490 GCAAGAGACTAAAGACTACCAGTGTGAAGATCTGCTTTCCAGGCTAAATCTGGAAG 1549

Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700  
Db 1550 TTCAATTCACAGCTCCACGTAGATAATAGTGGAAATCTTCGACAGGAGGATAGATGAG 1609  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 1610 AGCGATGATCTCGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCCAGACCT 1669  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 1670 CAGAGCTTCTGCCACTCTCCAGCATCTCCCGCCAGGAC-----AAGCTGGCACTG 1720  
Qy 741 AspalalaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 1721 CCAGATTTCAGCTCCAAAGGACAAACAGCGGCTC----- 1756  
Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProProAlaGlnProAla 777  
Db 1757 ---TCCTATGGAGCTTCCACCAACCATGATCTTCGTTTCCACCAAGCAGAGCTCGCCACA 1813  
Qy 778 AlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThr 797  
Db 1814 TCACCA---ACCACGAGGCTCCCTCTGCTCTAGGAACGCGGGAAGGTCCAACT 1870  
Qy 798 AlaSerSerAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArg 817  
Db 1871 GGCCCACTTCCACATCCCTCTAAGCACCCAGACCTCTAGT-----GGCAGCTCC 1921  
Qy 818 GlnArgSerSerAspProProAlaValHisProProLeuProProLeuArgValThr 837  
Db 1922 ACCCTATCCAAAGAGGCTCTCT-----CCCCCACCACCCGCGACACAAGAGAACC 1972  
Qy 838 SerThrAsnProLeuThrProThrPro----- 846  
Db 1973 CTATCCGACCTCCAGCCCACTACCTCATGGGCCCCCAACAAAGGCGCAGTTCTCTGG 2032  
Qy 847 -----ProProProValAlaLysThrProSerValMetGluAlaLeuSer 861  
Db 2033 GGTAAACGATGGGGTCCATCTCTTCAAGTAAGACTACAAACAAGTTTGAGGGACTATCC 2092  
Qy 862 Gln----- 862  
Db 2093 CAGCAGTCGAGCACCAGTTCTGCAAGAGACTGCCTTGGCCCCAAGAGTTCTTCTTAAACTA 2152  
Qy 863 ProSerLysProAla-----Pro 868  
Db 2153 CCTCAGAAAGTGGCACTAAGGAAACAGATCATCTCTCCTTAGACAAAGCCACCATCCCG 2212  
Qy 869 ProGlyIle-----SerGlnIle-----ArgProProPro--- 878  
Db 2213 CCCAAATCTTTCAAGAAATCATCAAGTTGGCAGAGTTGCCAANAAGCCACCACTGGA 2272  
Qy 879 ---LeuProProGlnProProSerArgLeuProGln----- 889  
Db 2273 GACCTGCCCCCAAGCCACAGAACTGSCCCCCCAAGCCCCCAATTTGGAGATTGGCGC 2332  
Qy 889 ----- 889  
Db 2333 CTAAGCCAGGAGAACTGCCCCCAACCAACAGCTGGGGAGCTGCCACCAACCCCAAC 2392  
Qy 889 ----- 889  
Db 2393 TCTCAGACTTACCTCCAAACCAACAGATGAAGGACCTGCCCCCCCAACCAACAGCTGGA 2452  
Qy 890 -----LysLysProAlaPro 894  
Db 2453 GACCTGTAGCAAAATCCCAGACTGGAGATGTCTCACCCCAAGGCTCAGCAACCTCTCAG 2512  
Qy 895 GlyThrAspLysSerThrProLeu-----Thr 903  
Db 2513 GTCACACTGAAGTACACCCCAATTTGGATCTATCCCCAAATGTGTCAGTCCAGAGACGCCATC 2572



```

Qy 904 AsnLysGlyGlnPro---ArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyPro 922
Db 2573 CAAAAGGAAGCATCTGGAGGACTCCCAACGACCTCAGCGCTACT----- 2617
Qy 923 LeuSerAsnAlaMetValLeuGlnProAlaProMetProArgLysSerGlnAlaThr 942
Db 2618 -----CTCCAGAGAGCGCCCTAGTACCTGCTCCAGAGAAATCAATACGGGG 2662
Qy 943 LysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGlu 962
Db 2663 AAAAAATAAGTGAGCGAGTGAAGACCATTTATGACTGCCAGGCAGACAACGATGACGAG 2722
Qy 963 LeuThrPheSerGluGlyAspValLeuValLeuValAspGlyGluGluAspGlnGluTyrTrp 982
Db 2723 CTCATTCATTCGAGGGAGAGTATTATCTGTCACAGGGGAAGAGGACCGAGGAGTGTGG 2782
Qy 983 IleGlyHisIleAspGly---AspPro-GlyArgLysGlyAlaPheProValSerPheVa 1001
Db 2783 ATTGCCCATCTGAAGGAGNAGCCCTTGAAGGAAGGGGTCTTTTCAGTGTCTTTGT 2842
Qy 1001 lHisPheIle 1004
Db 2843 TCATATCTGT 2852

```

## RESULT 4

```

US-10-104-047-981
; Sequence 981, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 981
; LENGTH: 3812
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-981

```

## Alignment Scores:

```

Pred. No.: 4,87e-148 Length: 3812
Score: 2089.50 Matches: 455
Percent Similarity: 59.20% Conservative: 153
Best Local Similarity: 44.30% Mismatches: 275
Query Match: 39.63% Indels: 145
DB: 17 Gaps: 19

```

US-09-914-042-1 (1-1006) x US-10-104-047-981 (1-3812)

```

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 47 ATGCCGAGCAGTTCAGCGTCCGCGGTCCTGCGCGTCACCGCGGAGGACCTCAGCTCC 106
Qy 21 ProThr---AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 39
Db 107 CCGGCTGGGGCGCGCGCTTCGCCGCCCAAGATGCCCGGTACCCGAGGGCGCGCTGGCG 166
Qy 40 IleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLys 59
Db 167 CGGGAGGAGATCTTGGAGAGGAGACCAAGCCATCTCGCAGAGAAATAAGAGGCTGTGGCG 226
Qy 60 AlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeu 79
Db 227 GCATTCATAGCTCCGGCTTGGCCATGTGGAGATGAAGCAGTACCGAGAGGCGCGTG 286
Qy 80 GluLysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeu 99
Db 287 GAATCTTTAGGCAACAGCACCCTGTGCCAGAACAGCCATGATGATGTCACAGGCTTCCCTA 346

```

```

Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMet 119
Db 347 AACTTGGCGCGTGTCCCGGAGGTTGCTCGCTCTTCAAGAACCTGATTTCAGAACTTG 406
Qy 120 AsnAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLys 139
Db 407 AACAACTATGTTCTCTTTCCCTTCGACAGTCTGATGAAGGGGCGAGTCAGGACGGTGA 466
Qy 140 GlyAspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLys 159
Db 467 CAGATTTCAAAAAACAAGCTGAGAGAGCATGGAAGGACTATGAAGCCAAATGGCCAG 526
Qy 160 IleGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179
Db 527 CTGGAAGAGGAG---CGCATCGGCGCAGGCTGACAGGAGGATC----- 568
Qy 180 GlyAlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCys 199
Db 569 CCTGGGAGGTGGGCCCAAGACATGCAGAGAGAGCGCGCATCTTCCAGCTGCACATGTGT 628
Qy 200 GluTyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219
Db 629 GAGTATCTGCTCAAAAGCGGGGAGAGCAGATGAAGCAAGTCTCTGACTTCTTCCAGCTGCACATGTGT 688
Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu 239
Db 689 CTCATCAAGTCTTCCCGCCAGCACAACTTTTTCGAAGATGGCTGGAAGGCTGCCCCAG 748
Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259
Db 749 AGCCTGTTCCCTTCATCGAAGCTGGCGCCTCAGTACATGCATCCATCAGGCCAG 808
Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
Db 809 GAGGACGAGCTACAGAAGCTGACCCAGCTCCGAGACTCCCTCCAGGAGACTGACGCTT 868
Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297
Db 869 GAGAGCAGAGAGGAACACCTTGAGCGGGAAGACTCAGGATGTGGCTATAGCATCCACCAG 928
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
Db 929 CACCAAGGCAACAAGCAGTTTGGGACGAGAAAGTGGGCTTTCTATACAAGAAAGTGCAC 988
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 989 GGAATTGAGAGTCTCGCAGAAAGGAAGTGTGGAGTCAAGTATGGCTGCTGACCATC 1048
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 1049 TCACACAGCAGATAAACCGCGCCCGGTGAAGCTGACCCCTGCTGACGTGCCAAGTAGG 1108
Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuLysSerHisAspArgThrTyrHisPhe 377
Db 1109 CCAAAACCTCAGGAGAAAAAGTGTTCGACCTGGTGCACCAACACCGGACGTACCACCTT 1168
Qy 378 GlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 1169 CAGGACAGGACGAGCAGAGTGTGAGCGGTGGGTGTGAGTGTTCAGTGTTCAGACAGCAAGGAC 1228
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAsp-----AspAsn 409
Db 1229 GAAGCCCTGAGCAGCGCTTCTCGGGGAGCCAGCGCTGGCCCGGGGTCTCGGGGGTCC 1288
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 1289 GCCCGCCATGATGGGAGCGGACGACCTCACAAAGTGTCTCATCGGCGGAGGTGAAGAGC 1348
Qy 430 MetThrGlyAsnAspValCysCysAspGlyAlaProAspProThrTrpLeuSerThr 449
Db 1349 AGGCTGGGAATACCCAGTGTCTGCGACTGCGGGGCTGCAGAGCCCAACCGTGTCTCAGCAC 1408

```

```
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 1409 AACCTGGCGGTGCTCACTTGATCGATGCTGGCGGTCCACCGGAACTGGCGGTGCC 1468
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 1469 TTTTCGGCATGTAGTCACTCACTTGACCTTGCTGGCCCTCGAGTGTGTGTCGCC 1528
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
Db 1529 TTGAACATGGGAAACACAGAGCTTCAATGAGGTATGGAGGGCCAGCTACCTCACACGGC 1588
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 1589 GGCCCTAAACCTCAGCTGAGAGTGACATGGGACCCCGAGGACTACATTATGGCCAAAG 1648
Qy 530 TyrIleGluArgTgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
Db 1649 TATGTGGAGCATAGTTTGCACGCCG-----TGCACACCTGAGCCTCAGCGA 1696
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
Db 1697 CTCTGGACAGCATTTGCAACAGGAGCCTCTGCTGCTACTGGAGGCCCTTTGCCAATGG 1756
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
Db 1757 CAGGACTTGGACAGCCCTGCCAGGCTGATGCACAGGACCTTGAAGAATCTGCTCTTG 1816
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
Db 1817 CATTGGCTGTCAAAAGTCGCCAACAGGCTCCCTGCTGCTGCTGTTGATTCATCATCCAG 1876
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
Db 1877 AACGGTGTCTACCTGGATGCCAAGCTGCTGCACGGGAAACAGCGCTCTGCATCAGCAGCA 1936
Qy 630 LeuThrAspAsnAlaGluCysLeuLeuLeuLeuArgGlyLysAlaSerIleGluIle 649
Db 1937 CTCTACACACGCGGAGTGTCTCAAGCTGTCTGAAGGGAGAGCTTTGTTGGGCACA 1996
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 1997 GTAAATGAAGCAGCGGACAGCAGCTCTGCACATAGCCAGGAAGACCAACAGGAGTGT 2056
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2057 GAGGAGCTGTGGAGCAGCGCCAGCGGGGACCTTTGCTCTTCCCTCTACATGTGCACTAC 2116
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2117 TCCTGGGTAAATTCACAGAGCTGGCTCTGACAGTGAGGAGATGAGGAAGAGAG--- 2173
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2173 ----- 2173
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2174 -----CGCTGCTTGTCTGAAGCTCCCGGCCAGGCT 2203
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
Db 2204 CACTGGGCCAGTGGGAGGCTGGACATCAGCAACAGACCTATGAGACTGTGCCAGCCTG 2263
Qy 769 GlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788
Db 2264 GGAGCA-----GCCACCCCTCAGGGCGAGAGTGAGGAGCTGTCCCCCGCCCTTGCCA 2314
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuThrLysThrAsn 808
Db 2315 GTCAAAAC-----TCTTCTCGGACTTTG----- 2338
Qy 809 SerValSerValAspGlyGlySerArgGlnArgSer-----Ser-SerAspProPr 825
```

```
Db 2339 -----GTCCAAAGGGTGTGCAAGACATCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2389
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2390 CTGAGTTTCAGAGGCCCTCGAGACCCCTGAGAGCCTGGGCAGTCCAG----- 2435
Qy 845 rProProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2436 -CCTCCTCTCTCACTGTATGAGCCCT-----TGGAACTTGGGGATCCCGACCA 2485
Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProProLeuProGlnProProSe 885
Db 2486 GCCCCACCACTCTGAAGAGGGCTCCGAGAGCCCGCAGCCTCCAGACCAGCCTG 2545
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLy 905
Db 2546 ACAT-----CCGGACCACTTCCGAGATGTACC----- 2576
Qy 905 sGlyGlnProArgGlyProValAspLysSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2577 -----TCCCGCTCAGAT---TCAGCTCCGAGACACTCGCTCTATCGCGG 2620
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 2621 GGGCGGGAGCCCTGAAGATGTCCTCAGCCAGG-CAGCTCTGCCAGAGGAAGCT 2679
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2680 GCCG----- 2683
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
Db 2683 ----- 2683
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2684 -----GTTGGCATCACTGAGGAGATGGCTCAAGACTGGGAGTCTCCAGCAGCTTC 2736
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2737 TGTGCAACTTTTGCAGAC 2755
RESULT 5
US-10-490-605-1
; Sequence 1, Application US/10490605
; Publication No. US20050019768A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCE, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; TITLE OF INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATL/SEN-A0121P-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; CURRENT FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4050
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-490-605-1
Alignment Scores:
Pred. No.: 5,26e-148 Length: 4050
Score: 2089.50 Matches: 455
Percent Similarity: 59.20% Conservative: 153
Best Local Similarity: 44.30% Mismatches: 275
Query Match: 39.63% Indels: 145
```

DB:	21	Gaps:	19
US-09-914-042-1 (1-1006) x US-10-490-605-1 (1-4050)			
Qy	1	MetProaspGlnIleSerValserGluPheValalaGluThrHisGluAspTyrLysAala	20
Dd	46	ATGCCGGAGCAGTTTCAGCGTGCCGAGTTCTGTGCCGTACCGCGGAGGACCCTCAGCTCC	105
Qy	21	ProThr---AlaSerSerPheThrThraArgThrAlaGlnCysArGsnThrValalaAala	39
Dd	106	CCGCTGTGGCGCGCGCTTCGCCGCCAAGATGCCCCGTACCCGAGGGCGCGCTGGCG	165
Qy	40	IledgluAlaLeuAepValasPargMetValleuTyrLysMetLysLysSerVallys	59
Dd	166	CGGAGAGAGATCTTGGAAAGAGCAACAACCATCTCTGCAGAGAATAAAGAAGCGTTGCGG	225
Qy	60	AlaIleAenSerSerGlyLeuAlaHisValGluAenGluGluGlnTyrThrGluAlaleu	79
Dd	226	GCAATCCATAGCTTCGGCTTGGCCATGTGGAGATGAAGACGAGTACCCGAGAGCGCGTG	285
Qy	80	GluLysPheGlyAsnCyValcysArgAspPProaspLeuGlySerAlaPheLeu	99
Dd	286	GAATCTTAGSCAACAGCACCTGTCCCAGAACGCCATGAGCTGCCACAGGCTTCCCTA	345
Qy	100	LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuIleGlnAenMet	119
Dd	346	AACVTGGCCGTGTTACCGCGGAGGTGTCTGCGCTCTTCAAGAACCTTGATTTCAGAACTT	405
Qy	120	AsnAenIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyVallys	139
Dd	406	AACAACATTGTCTCTTCCCTCGACAGCTGTGATGAAGGGCGAGCTGAGGAGCGGTGCA	465
Qy	140	GlyAspLeuLysLysProPheAspLysAlatrpLysAspTyrGluThrLysIleThrLys	159
Dd	466	CAGGATTCCAAACACACTGGAGNAGGCATGGAAGGACTATGAAGCCAAAATGCCCNAG	525
Qy	160	IleGluLysGluLysLysGluHiAlalalsLeuHisGlyMetIleAthrGluIleSer	179
Dd	526	CTTGAGAGAGAG--CGCGATCGGCGCCAGGGGTGACAGAGGGATC-----	567
Qy	180	GlyAlaGluIleAlaGluMetGluLysGluHargPhePheGlnLeuGlnMetCys	199
Dd	568	CCTGGGAGGTGGCCAGGACATGCAGAGAGCGGCGCATCTTCCAGCTGCACATGTGT	627
Qy	200	GluTyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAen	219
Dd	628	GAGTATCTGCTCAAGCCGGGAGAGCCAGATGAAGCAAGTCTGACTTCTTCCTTCAGAGC	687
Qy	220	LeulleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu	239
Dd	688	CTCATCAAGTTCTTCCACGCCCCAGACAACCTTTTTCCAAGATGCTGGNAGGCTGCCAG	747
Qy	240	SerLeuLysProSerilleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln	259
Dd	748	AGCCTGTCCCCTTCATCGAAGACTGGCGGCTTCAGTACATGCACCTCCATCAGGCCCCAG	807
Qy	260	AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal	279
Dd	808	GAGACAGAGCTACAGAGACTGACCAGCTCCGGAGCTCCCTCCGAGGGACACTGAGGCTT	867
Qy	280	GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln	297
Dd	868	GAGAGCAGAGGAACACCTCAGCCGGAGAAGAACTCAGGATGTGGCTATAGCATCCACCAG	927
Qy	298	ProGlnGlyAanLysGluHisGlyThrCluarAenGlySerLeuTyrLysLysSerAsp	317
Dd	928	CACCAAGGCCAACAGACAGTTTGGGACGGAGAAGTGGGCTTCTATATACAGAAAAAGTGAC	987
Qy	318	GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAenGlyPheLeuThrIle	337
Dd	988	GGAATTCGAAGAGTCTGCACGAAGGAGTGTGGAGTCAAGTATGGCTGCCTGCACCATC	1047
Qy	338	SerHisGlyThrAlaAenArgProAlalalsLeuAenLeuLeuThrCysGlnVallys	357

```
Db 2116 TCCTGGGTAATTTCCACAGAGCCTGGCTCTGCAGCTGAGGAGGATGAGGAAGAGAAAG--- 2172
Qy 710 GlnProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2172 -----
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2173 -----
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
Db 2203 CACTGGGCGCAGTGGGAGGCTGGACATCAGCAACAGACTATGAGACTGTGCCACGCTG 2262
Qy 769 GlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788
Db 2263 GGAGCA-----GCCACCCCTCAGGCGGAGAGTGAGGACTGTCCCCGCCCTTGCCA 2313
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsn 808
Db 2314 GTCAAAAC-----TCTTCTCGGACTTGG----- 2337
Qy 809 SerValSerValAspGlyGlySerArgGlnArgSer-----Ser-SerAspProPr 825
Db 2338 -----GTCCAAGGGTGTCAAGACATGCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2388
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2389 CTGAGTTCCAGAGGCCCTCAGACCCCTCAGAGCCTGGCGAGTCCAG----- 2434
Qy 845 rProProProProAlaAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLy 865
Db 2435 -CCTCTCTCTCAGTCTGATGAGCCCT-----TGGAACTTGGGGATCCAGCCAA 2484
Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProLeuProProGlnProProSe 885
Db 2485 GCCCCACCACTCTGAAGAGGGCTCCGAGAGCCCCCAGGACCTCCAGACCCAGCCTG 2544
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLy 905
Db 2545 ACAT-----CCGGGACCACTCCGAGAGATGTACC----- 2575
Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2576 -----TCCCGTCTCAGAT---TCAGTCCGAGAGCACTCGTCTCTATCGGCGG 2619
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerGl 940
Db 2620 GGGGCGGAGGCGCTGAAGATGGTCCCTCAGCCAGG-CAGCCTCTGCCCAGAAGGAAGT 2678
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2679 GCCG----- 2682
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGl 980
Db 2682 ----- 2682
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2683 -----GTTGGCATCACTGAAGGAGATGGCTCAAGGACTGGGAGTCTCCACGAAGTTC 2735
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2736 TGTGCAACTTTTGCAAGAC 2754
```

## RESULT 6

```
US-09-764-875-37
; Sequence 37, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
```

```
; FILE REFERENCE: PUZ02
; CURRENT APPLICATION NUMBER: US/09/764.875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 3436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-37

Alignment Scores:
Pred. No.: 2,82e-126 Length: 3436
Score: 1801.50 Matches: 346
Percent Similarity: 87.85% Conservative: 1
Best Local Similarity: 87.59% Mismatches: 3
Query Match: 34.16% Indels: 45
DB: 11 Gaps: 1

US-09-914-042-1 (1-1006) x US-09-764-875-37 (1-3436)

Qy 612 GlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThr 631
Db 1 GGGAACTCTGGATAAACAGACAGAGGAGGAAAGGCAGCACAGCCCTGCACCTACTGCTGCTGACC 60
Qy 632 AspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsn 651
Db 61 GACAATCCGAGTCCCTCAAGTTGCTCTCTGGGGGAGGCGCTCCATCGAGATAGCAAAAC 120
Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671
Db 121 GAGTCAGAGAGAGACTCCGCTGGACATTGCCAAGCGCCTCAAGCAGCAGCACTGTGAGGAG 180
Qy 672 LeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691
Db 181 CTGCTGACCCCAAGCCTTATCTGGAAGATTTAATTTCTCACGTTTACGTTGAATATGAATGG 240
Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711
Db 241 CGACTACTCCACGAAAGACCTGGATGAAGATGATCAGCAGATGGATGAGAAATTTGAGCGCC 300
Qy 712 SerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeu 731
Db 301 AGTCCCAACCGCGGGAGACCGGCCCATCAGCTTCTTACCAGCTGGGCTCCCAACAGCTT 360
Qy 732 GlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArg 751
Db 361 CAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTTGCCAAGGAGAGCAGAGG 420
Qy 752 AlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPro 771
Db 421 GCTTTCATGCTCCAGCATCTTCGAGAATGAGACTTACGGAGGCCCTCTCAGTGGAGGCCCA 480
Qy 772 ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsn 791
Db 481 CCTCCGCCCAAGCTCGAGCCGCCAGCACCCAGCGCCCCCGCTTCTCCACGGAAT 540
Qy 792 ValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSer 811
Db 541 GTTGGCAAA----- 549
Qy 812 ValAspGlyGlySerArgGlnArgSerSerAspProProAlaValHisProProLeu 831
Db 549 ----- 549
Qy 832 ProProLeuArgValThrSerThrAsnProLeuThrProThrProProProValAla 851
Db 550 -----GATCCCTGACCCCGCCAGCGGSCCCAGCCCGTTCGCC 585
Qy 852 LysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIle 871
Db 586 AAGACGCCGCGTAATGGAAGCCTTGAGCCAGCGGAGCAAGCCTGCCCCCGCTGGGATC 645
```

QY 872 SerGlnIleArgProProLeuProProGlnProProSerArgLeuProGlnIleLys 891  
 DB 646 TCACAGATCAGCGCCCACTCTGCCCCACAGCGCCGCTCCCGCAGAAGAAG 705  
 QY 892 ProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyPro 911  
 DB 706 CTGTGCGCGGGGCTGACAGTCCACCCACTGACCAACAAGGCCAACCCGAGAGGACCT 765  
 QY 912 ValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnPro 931  
 DB 766 GTGGATCTCTGCAACGGAAGCTCTGGTCTCTGTCTCAATGCTATGCTCTGCAGGCC 825  
 QY 932 ProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAla 951  
 DB 836 CCTGCACCCATGCTAGGAGTGCAGGCAACCAAGTTGAAGCCCTTAAGCGGGTGAAGCG 885  
 QY 952 LeuThrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIle 971  
 DB 896 CTCTATAACTGTGTGGCTGACAAACCCCGATGACCTCTTCTCCGAGGGGGATGTGATC 945  
 QY 972 IleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGly 991  
 DB 946 ATCGTGGACGGGAGGAGCACGAGGTGTGGATTGGCCACATTTGATGGAGATCCTGGT 1005  
 QY 992 ArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006  
 DB 1006 CGCAAGCGCATTCCTCGGTGTCATTTGTGCATTTATCGCTGAC 1050

RESULT 7  
 US-09-764-875-358  
 ; Sequence 358, Application US/09764875  
 ; Publication No. US20040018969A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P0202  
 ; CURRENT APPLICATION NUMBER: US/09/764,875  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1249  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 358  
 ; LENGTH: 3413  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-875-358

Alignment Scores:  
 Pred. No.: 4,26e-123 Length: 3413  
 Score: 1759.50 Matches: 338  
 Percent Similarity: 86.48% Conservative: 1  
 Best Local Similarity: 86.22% Mismatches: 8  
 Query Match: 33.37% Indels: 45  
 DB: 11 Gaps: 1

US-09-914-042-1 (1-1006) x US-09-764-875-358 (1-3413)

QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634  
 DB 8 GATTAACAGACAGGGAAGGAGCAGCAGCCCTGCACTACTGCTGCTTACCGACAATGCC 67  
 QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654  
 DB 68 GAGTGCCTCAAGTTGTYCTCTGGGGGGAAGGCGCTCCATCGAGATAGCAACGATCAGGA 127  
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuThr 674  
 DB 128 GAGACTCCGCTGGACATTTGCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGTYGTGACC 187  
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeu 694  
 DB 188 CAAGCTTATTTGGAAGATTTTAAATTTCTCACGTTTCACTGTTGAATGAATGCGGACTACTC 247

QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714  
 DB 248 CACGAAGACCTGGATGAAAGTGATGACGACATGATGAGAAATTGACGCCAGTCCCAAC 307  
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734  
 DB 308 CGGCGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCAGTCTAAC 367  
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754  
 DB 368 GCTGTATCTTTGGCCAGAGATGCTGCAAACTTTCGAAGGAGAGAGCAGAGGGCTTCATG 427  
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774  
 DB 428 CCCAGCATCTTGCAGAAATGAGCTTACGAGCCCTCTCTAGTGGCAGGCCACCTCCGCC 487  
 QY 775 GlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794  
 DB 488 CAGCTCGAGCCCGCAGCACCCAGCGCCCTTCTCCACGAAATGTTGGCAAA 547  
 QY 795 ValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGly 814  
 DB 547 ----- 547  
 QY 815 GlySerArgGlnArgSerSerAspProProAlaValHisProProLeuProProLeu 834  
 DB 547 ----- 547  
 QY 835 ArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPro 854  
 DB 548 -----GATCCCTTCAGCCCAACGCGGCGGCCCGCCCGTTCGCAAGACGCC 592  
 QY 855 SerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerGlnIle 874  
 DB 593 AGCGTATGAGAGCTTGGACGACCGCAGCAGGCTGCCCCCTGGGATTTCCACAGATC 652  
 QY 875 ArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPro 894  
 DB 653 AGGCCCCCACCCTCTGCCCCACAGCGCCCGCCAGCGCTCCCGCAGAGAGCCTGCGCG 712  
 QY 895 GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeu 914  
 DB 713 GGGGCTGACAAAGTCCACCCCACTGACCAACAAAGGCGCAACGAGAGGACCTGTGGATCTC 772  
 QY 915 SerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPro 934  
 DB 773 TCTGCAACGGAGCTCTGGGTCTCTGTCCAAATGCTRTGGTCTCTGCAGCCCCCTGCACCC 832  
 QY 935 MetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuThrAsn 954  
 DB 833 ATGCTAGGAAGTCCGAGGCAACCAAGTTGAAGCTTAAGCGGTGAAGAGCGCTCTATAAC 892  
 QY 955 CysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAsp 974  
 DB 893 TGTGTGGCTGACAAACCCCGATGACCTCACCTTCTCCGAGGGGGATGTGATCATCTGGAC 952  
 QY 975 GlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgLysGly 994  
 DB 953 GGGGAGGAGGAGGAGGAGTGTGATTGGCCACATTTGATGGAGATCTCTGTGGCAAGGC 1012  
 QY 995 AlaPheProValSerPheValHisPheIleAlaAsp 1006  
 DB 1013 GCATTCCTCGGTGTCATTTGTGCATTTTATCGCTGAC 1048

## RESULT 8

US-10-104-047-314  
 ; Sequence 314, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
 ; FILE REFERENCE: H1-A0105

```
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 314
; LENGTH: 3123
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-314

Alignment Scores:
Pred. No.: 6 72e-89 Length: 3123
Score: 1307.50 Matches: 253
Percent Similarity: 84.3% Conservative: 1
Best Local Similarity: 84.05% Mismatches: 2
Query Match: 24.80% Indels: 45
DB: 17 Gaps: 1

US-09-914-042-1 (1-1006) x US-10-104-047-314 (1-3123)

Qy 706 AspGluLysLeuGlnProSerGluAsnArgGluAspArgProIleSerPheTyrGln 725
Db 2 GATGAGAAATTCAGCCAGTCCCAACCGGGGGAAGACCGGCCCATCAGCTTCTACCA 61
Qy 726 LeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAsnLeu 745
Db 62 CTGGGCTCCAAACACGCTTCAGTCTAACGCTGTATCTTTGGCAGAGATGCTGCAAC 121
Qy 746 AlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAla 765
Db 122 GCCAAGGAGAGACGAGGGCTTTCATGCCAGCATCTTGCAGATGAGACTTACGAGCC 181
Qy 766 LeuLeuSerGlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaPro 785
Db 182 CTCCTGAGTGGCAGCGCCACCTCCCGCCAGCCTGCAGCCCGCCAGCACCCAGCGCC 241
Qy 786 ProLeuProProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrp 805
Db 242 CCGCTTCCTCCACGGAATGTGGCAAA----- 268
Qy 806 LysThrAsnSerValSerValAspGlyLysSerArgGlnArgSerSerSerAspPro 825
Db 268 ----- 268
Qy 826 AlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProThr 845
Db 269 -----GATCCCTGACCCCCACG 286
Qy 846 ProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 287 CGGCCCCACCGTGTGCCAAGACGCCAGCGTAATGGAGCCTTGAGCAGCCGAGCAAG 346
Qy 866 ProAlaProProGlyLysSerGlnIleArgProProProLeuProProGlnProProSer 885
Db 347 CCGTCCCGCCCTGGGATCTCAGATCAGGCCCCCACTCTGCCCCACGCGCCACG 406
Qy 886 ArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 407 CGCCTCCGCGAAGAAGCCTCGCGCGGGGTGACAAGTCCACCCCACTGACCAACAAA 466
Qy 906 GlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsn 925
Db 467 GGCACCAACGAGAGGACCTGTGGATCTCTGCAACGGAAGCTCTGGGTCTCTGTCCAAT 526
Qy 926 AlaMetValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLys 945
Db 527 GCTATGGTCTCGACGCCCTGCACCCATGCCTAGGAAGTGCAGGCAACCAAGTTGAAG 586
Qy 946 ProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPhe 965
Db 587 CCTAAGCGGGTGAAGCGCTCTATAACTGTGTGGCTGACAAACCCCGATGAGCTCACCTTC 646

Qy 966 SerGluGlyAspValIleIleValAspGlyGluGluAspGlnGluTyrTrpIleGlyHis 985
Db 647 TCCGAGGGGATGATCATCTCTGACGGGAGGAGGACCCAGGAGTGGTGGATTGGCCAC 706
Qy 986 IleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAla 1005
Db 707 ATTGATGGAGATCTCTGTCGCAAGGCGCATTCCTCCGGTGTCTATTTGTGCACTTTAT 766

Qy 1006 Asp 1006
Db 767 GAC 769

RESULT 9
US-09-815-343-1154
; Sequence 1154, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1154
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(563)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-1154

Alignment Scores:
Pred. No.: 1.95e-43 Length: 563
Score: 695.00 Matches: 144
Percent Similarity: 82.12% Conservative: 3
Best Local Similarity: 80.45% Mismatches: 14
Query Match: 13.18% Indels: 18
DB: 9 Gaps: 1

US-09-914-042-1 (1-1006) x US-09-815-343-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 2 AACATGCAATTTAAGGGGATGACAATACATCGAGAAAATAACATCGTCCAAGAACTGACA 61
Qy 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 62 AAGGAGATCATCTCAGAAAGTCTCAGGATGACGGGCAATGACGTCTGCTGTGACTGTGG 121
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 122 CGCCAGATCTCATAGCTTTTCCACCAACCTGGGCACTCCAGGATGACCTCGAGTGTTC 181
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 182 GGAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGGATGACCTCCCTGACCTTAGATGA 241
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 242 CTGGGAACATCTGAGCTGCTCTGCCCAAGAAATATTGGGAATGCGAGGCTTTAATGAGATC 301
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMetAs 520
Db 302 ATGGAATGTTCCCTACAGCTGAGACTCAGTCAAAACCAACCCAGGCGCGGCACATGAA 361
Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg-ArgTyr-AlaArgLysLys 539
```

```

Db 362 TGCAGAAAGGACTACATCAGCCCAAGTACATCAGAGGANGATACCCCAAGGAAG 421
Qy 540 His-AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAsp11 559
Db 422 CACCCCGGATACGCGCGCGA----- 443
Qy 559 ePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLys 575
Db 444 ----AANCTTTCCAGTCTTTTGNNAAGGCGNGTCCAAACCNAAG 488

RESULT 10
US-10-097-105-1154
; Sequence 1154, Application US/10097105
; Publication No. US20040037842A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Segrist, Heather
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.504C1
; CURRENT APPLICATION NUMBER: US/10/097,105
; CURRENT FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 1562
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1154
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 403, 446, 464, 465, 472, 483, 485, 499, 539, 554
; OTHER INFORMATION: n = A,T,C or G

US-10-097-105-1154
Alignment Scores:
Pred. No.: 1,95e-43 Length: 563
Score: 695.00 Matches: 144
Percent Similarity: 82.12% Conservative: 3
Best Local Similarity: 80.45% Mismatches: 14
Query Match: 13.18% Indels: 18
DB: 18 Gaps: 1

US-09-914-042-1 (1-1006) x US-10-097-105-1154 (1-563)
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 2 AACATGCAATTAAGGGGGATGACAATACATCTGGAGAAAAATAACATCGTCCAAGAACTGACA 61
Qy 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGly 440
Db 62 AAGGAGATCATCTCAGAAGTCTGAGGATGACGGCAATGACGTCTGCTGACTGTGGG 121
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 122 GCGCCAGATCCTACATGCTTTCCACCAACCTGGGCATCCTGACCTGATCGAGTGTCC 181
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 182 GGAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGGATGACGCTCCCTGACCTTAGATGA 241
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 242 CTGGGAACATCTGAGCTGCTGCCCAAGAAATATGGGAATGCAAGGCTTTAATGAGATC 301
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMetAs 520
Db 302 ATGGAATGTTGCTTACAGCTGAGGACTCAGTCAACACCCACCCAGGCGGACATGAA 361
Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg-ArgTyr-AlaArgLysLys 539

```

```

Db 362 TGCAGAAAGGACTACATCAGCCCAAGTACATCAGAGGANGATACCCCAAGGAAG 421
Qy 540 His-AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAsp11 559
Db 422 CACCCCGGATACGCGCGCGA----- 443
Qy 559 ePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLys 575
Db 444 ----AANCTTTCCAGTCTTTTGNNAAGGCGNGTCCAAACCNAAG 488

RESULT 11
US-10-037-270-155
; Sequence 155, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghaast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 155
; LENGTH: 6990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)..(2397)

US-10-037-270-155
Alignment Scores:
Pred. No.: 1.09e-32 Length: 6990
Score: 571.50 Matches: 196
Percent Similarity: 41.68% Conservative: 137
Best Local Similarity: 24.53% Mismatches: 305
Query Match: 10.84% Indels: 161
DB: 15 Gaps: 25

US-09-914-042-1 (1-1006) x US-10-037-270-155 (1-6990)
Qy 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyrLysMetLysLys 56
Db 115 GCAGCTTTGGGAAGAGTAGAAGGTGTGCGCAATTTGGAACATA---AAACTTGATGANG 171
Qy 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThr 76
Db 172 CTTGTGAAACTT-----GGGATTGCAATGATTGATCTGGAAGAAAGCCTTTTGT 219
Qy 77 GlnAlaLeuGluLysPheGlyGlyAsnCysValCysArgAsp----- 90

```

Db 220 GTTGCATAAATAACAGTTCATGAATGGG-----ATTGAGAGACCTGGCCAGTATTCTAGT 273  
 Qy 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109  
 Db 274 AATGATGCTGCTGTGAGACAAGTTTGACCAAGTTTCTGACAGTCTTCAAGAAATGATA 333  
 Qy 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSer 129  
 Db 334 AATTTTCACACATCTCTGTGCCAAACTCAGAGATCAATTTAAGCCACAGCTTCAGTAAC 393  
 Qy 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPheAspLysAla 149  
 Db 394 TTGTGTTAAGAAGATCTTAGAAAATTCAAA---GATGCCAAGAAGCAATTTGAAAAGTC 450  
 Qy 150 TrpLysAspTyrGluThrLysIleThrLys-----IleGluLysGluLysLysGlu 166  
 Db 451 AGTGAAGAAAAGAAAATGCGTTAGTAAAAAATGCCCAAGTACAAAGAAACAAACACAT 510  
 Qy 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGlu 186  
 Db 511 GAAGTTGAA-----GAAGCCACCACACATT 534  
 Qy 187 MetGluLysGluArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsn 206  
 Db 535 CTGACAGCAACAAGAAAATGTTCCGACACATAGCCCTCGATTATGTGCTTCAGATTAA 594  
 Qy 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAla 226  
 Db 595 GTTCTTCAATCAAAAAGAGATCAGAAAATCCTAAATCAATGTTGTCTATTATGTATGCC 654  
 Qy 227 GlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGlu 246  
 Db 655 CATTGGCTCTCTTCATCAGGATATGATCTGTTAGTGAACCTTGGACCTTACATGAAG 714  
 Qy 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgArgGlnLeu 266  
 Db 715 GATCTTGTGTGCACAGTGTGATCGACTGTTGTGGATGCGCAAGAGGAGAAAGAGAAATG 774  
 Qy 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluLysGluAspSerGln 286  
 Db 775 GAGCAAAACATTC-----ACCATTCAACAAAGAGATTTCTCCAGT 816  
 Qy 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306  
 Db 817 GATGATTTCAAGTTAGATATAACGTA-----GATGCTGCAAAATGGCATA 861  
 Qy 307 GluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArg 326  
 Db 862 GTTATGGAAGGATATCTGTTCAACAGCAGCCAGCAATGCTTCAAAAATTTGGAACAGGCGC 921  
 Qy 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346  
 Db 922 TGGTTTTCAATACAGAAATACAGTTGTTTACCAGAAAAAATTTAAGGATAATCCGACT 981  
 Qy 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGluLys 363  
 Db 982 GTGTAGTTGAAGACCTCAGGCTTTGCACGTGAAAAACATTTGGAAGACATAGAGCGCAGA 1041  
 Qy 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383  
 Db 1042 TTCTGCTTTGAGTGGTCTCGCCAAACAAAAAGTTGCTCATGCTCCAGCGAGATTTCCGAAAAG 1101  
 Qy 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAla 403  
 Db 1102 CTGCGCCAGGATGGAATTAAAGCTCTTCAGACCAGTATTGCTTACTGCTTATAGAGAG--- 1158  
 Qy 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412  
 Db 1159 ---AAGGTGATGATCAGAGAGCTGGTAAGAAATCATCTCCATCCACAGGAAGCCTA 1215  
 Qy 413 -----AsnAsnIleValGlnGluLeuThrLys-----GluIleIleSerGluVal 427  
 Db 1216 GATTCTGGAATAGTCCAAAGAGAAATTTATGAAAAGGAGAAAGTGCGCTTCAGCGGCTC 1275

Qy 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447  
 Db 1276 CAGTGTATCCCTGCCAATGCCAGCTGTGTGACGTGTGGCCTGGCAGATCCACGGTGGGCC 1335  
 Qy 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467  
 Db 1336 AGCATCAACCTGGGCATCACCTGTGTATCGAGTGTCTCCGAATTCACCGGAGCTTTGGG 1395  
 Qy 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487  
 Db 1396 GTTCATTTTCAAAAGTACGATCTTTAACTTTAGACACCTGGAGCCAGAACTTTTAAAG 1455  
 Qy 488 LeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAla 507  
 Db 1456 CTTATGTGTGAGTTGGGGAATGATGTTATAATCGAGTTTATGAAGCTAATGTGAAAAA 1515  
 Qy 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527  
 Db 1516 ATGGGAATAAAGAAACCCCAACACAGGA---CAAAGACAGGAGAGGACATATATCAGA 1572  
 Qy 528 AlaLysTyrIleGluArgArgTyrAla----- 536  
 Db 1573 GCAAAATATGTGGAGAGAGAAATTTGTGGATAAAATATTCTATATCATTTATCATCTCCTGAG 1632  
 Qy 537 ---ArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCys----- 551  
 Db 1633 CAGCAAAAAAAGTTGTCTCTAAAGTTCTGAAGAAAGAGGCTGAGCATTTCTAAATTT 1692  
 Qy 552 -----GluAlaValLysThrArgAspIlePhe 560  
 Db 1693 GGGCCAGGGGACCAAGTCAGAGCATCTGCCAAAGTTTCAGTCAGAGTAATGAC---AGT 1749  
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
 Db 1750 GGAATTCAGCAGAGCTCTGATGATGGAAGAAATCTTTACCTCCAGGCTGTAGCCAAT 1809  
 Qy 581 Gly---HisGluProAsp----- 585  
 Db 1810 AGTTATATGAGCTGAAGAGAAAGGCAAGATTTCTTATGTTTCTTGACTCGAAACAT 1869  
 Qy 585 ----- 585  
 Db 1870 CTTAATCCAGGACTTCAGCTTTATAGGCGCTCATATGAAAAAACCTTCTCTAAATGGCT 1929  
 Qy 586 GluThrAlaLeuHisLeuAla----- 592  
 Db 1930 GAGCTTTGGCTCATGTGTCAGACGCTGAACCTGGGCCAATTCAGAGGAAAAACAAAGCGACA 1989  
 Qy 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610  
 Db 1990 CCATTTATTCAGGCTGTATTAGGGGCTCTTTGTGTGAGCTGTGAGTTCTCTCTACAGAA 2049  
 Qy 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630  
 Db 2050 GGTCTAATGTCAACCAAGAGATGTCCAAGGGCGGGACCATTTGCACCATGCCACCGTC 2109  
 Qy 631 ThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAla 650  
 Db 2110 TTAGGCACACAGCGCAGGTATGTTTATTCCTAAACAGAGGTGCCAATCAACATGCCACT 2169  
 Qy 651 AsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670  
 Db 2170 GATGAGAGAGGAAGACCTTTTGAGCATAGCTGTGGAGCAGGCAATGCTGATATAGTC 2229  
 Qy 671 GluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690  
 Db 2230 ACCTTGTTACGTTTAGCA----- 2247  
 Qy 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710  
 Db 2248 -----AGAATGAATGAAGATGCGGGAATCAGAAAGACTTTATGGA-----CAG 2292



QY 711 ProSerGluAsnArgArgGluAspArgProIleSerPheThrGlnLeuGlySerAsn 729  
 Db 2293 CCAGGTGTAAGAACTTATCAGGACATATTCGTGATTTTCCCAATATGGCATCCAAT 2349

## RESULT 12

US-10-117-722-155  
 ; Sequence 155, Application US/10117722  
 ; Publication No. US20030219744A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Asundi, Vinod  
 ; APPLICANT: Zhang, Jie  
 ; APPLICANT: Drmanac, Radotje T.  
 ; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
 ; FILE OF INVENTION: Polypeptides  
 ; FILE REFERENCE: 784CIP2BCIP  
 ; CURRENT APPLICATION NUMBER: US/10/117,722  
 ; PENDING FILING DATE: 2002-04-04  
 ; PRIOR APPLICATION NUMBER: 09/620,312  
 ; PRIOR FILING DATE: 2000-07-19  
 ; PRIOR APPLICATION NUMBER: 09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; NUMBER OF SEQ ID NOS: 1104  
 ; SOFTWARE: pt\_FL\_genes Version 1.0  
 ; SEQ ID NO 155  
 ; LENGTH: 6990  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (61)..(2397)  
 US-10-117-722-155

## Alignment Scores:

Pred. No.: 1.09e-32 Length: 6990  
 Score: 571.50 Matches: 196  
 Percent Similarity: 41.68% Conservative: 137  
 Best Local Similarity: 24.53% Mismatches: 305  
 Query Match: 10.84% Indels: 161  
 DB: 17 Gaps: 25

US-09-914-042-1 (1-1006) x US-10-117-722-155 (1-6990)

QY 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyrLysMetLysLys 56  
 Db 115 GCAGCTTTGGAGAAAGTAGAGGTGATGGCAGAAATTTGGAACTA---AAACTTGATAAG 171  
 QY 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThr 76  
 Db 172 CTTGTGAACCTT-----GGGATTGCAATGATTGATGATCTGGAAGAGCCCTTTGT 219  
 QY 77 GlnAlaLeuGluLysPheGlyCysValCysArgAsp----- 90  
 Db 220 GTTGCAATAAACAAGTTCATGAATGGG-----ATTGAGACCTGGCCAGTATTCTAGT 273  
 QY 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109  
 Db 274 AATGATGCTGCTGGAGACAAAGTTTGACCAAGTTTTCTGACAGTCTTCAAGAAATGATA 333  
 QY 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnIleIleSerPheProLeuAspSer 129  
 Db 334 AATTTTCACACATCTGTTGCCAACTCAGAGATCAATTTAAGCCAGCTTCAGTAAC 393  
 QY 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysPheAspLysAla 149  
 Db 394 TTTGTTAAAGAAGATCTTAGAAAATTCAA---GATGCCAAGAACAATTTGAAAAGTC 450  
 QY 150 TrpLysAspTyrGluThrLysIleThrLys-----IleGluLysGluLysLysGlu 166  
 Db 451 AGTGAAGAAAAGAAAATGCGCTTAGTAAAAAATGCCAAGTACAAAGAAACAAACAT 510

QY 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGlu 186  
 Db 511 GAAGTTGAA-----GAAGCCACCAACAT 534  
 QY 187 MetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsn 206  
 Db 535 CTGACAGCAACAGAAATGTTCCGACACATAGCCCTCGATATATGTCTTATGATGAT 594  
 QY 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAla 226  
 Db 595 GTTCTTCAATCAAAAAGGAGATCAGAAATCTTAAATCAATGTTGTTTATGATGCC 654  
 QY 227 GlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysPheSerIleGlu 246  
 Db 655 CATTGGCTTCTTTCATCAAGGATATGATCTGTTAGTGAATCTGGACCTTACATGAAG 714  
 QY 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeu 266  
 Db 715 GATCTTGGTGACAGTTGGATCGACTGTTGTCGATGTCAGCAAGAGAGAAATG 774  
 QY 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlnLysGluAspSerGln 286  
 Db 775 GAGCAAAACATTC-----ACCATTCAACAAAGATTTCTCCAGT 816  
 QY 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306  
 Db 817 GATGATTTCTAAGTTAGAAATAAAGCTA-----GATGCTGCAATGGCATA 861  
 QY 307 GluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArg 326  
 Db 862 GTTATGGAAGATATCTGTTCAACAGCCAGCAATGCTTCAAACTTTGGAAGGCGC 921  
 QY 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346  
 Db 922 TGGTTTTCATACAGATATATCAGTTGTTTACCAAAAAATTTAAGATATCCGACT 981  
 QY 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluLys 363  
 Db 982 GTGGTAGTTGAAGACCTCAGGCTTTGCACAGTGAACATTTGTGAAGACATAGAGCCAG 1041  
 QY 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGln 383  
 Db 1042 TTCTGCTTTGAGGTGCTCGCCCAACAAAAATTTGCATGCTCCAGGAGATTTCCGAAA 1101  
 QY 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403  
 Db 1102 CTGGCCAGGATGATTAAGCTGTTTCAGACCATGTTGCTACTGCTTATAGAGAG--- 1158  
 QY 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412  
 Db 1159 ---AAGGTGATGAATCAGAGAAGCTGTAAGAAATCATCTCCATCCACAGGAAGCCTA 1215  
 QY 413 -----AsnAsnIleValGlnGluLeuThrLys-----GluIleIleSerGluVal 427  
 Db 1216 GATTCTGGAATAGTCCAAAGAGAAATTTTGAAGAGGAGAAAGTGCCTTTCAGGGGTC 1275  
 QY 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447  
 Db 1276 CAGTGTATCCCTGCAATGCCAGCTGTTGTGACTGTGCTGCGGATGTCAGGTTGGGCC 1335  
 QY 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467  
 Db 1336 AGCATCAACCTGGGCATCACCTTGTGATCGAGTCTCCGGAATTCACGGAGCCTTGG 1395  
 QY 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487  
 Db 1396 GTTCATTTTCAAAGTACGATCTTAACTTTAGACACCTGGGAGCCAGAACTTTTAAAG 1455  
 QY 488 LeuAlaLysAsnIleGlyAspAlaGlyPheAsnGluIleMetGluCysLeuProAla 507  
 Db 1456 CTTATGTGTGAGTTGGGGAATGATGTTATAAATCGAGTTTATGAGCTTAATGTGGAANA 1515

Qy 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527  
 Db 1516 ATGGGAATAAGAAACCCCAACGAGA---CAAGACAGGAGGAGGCGATATATACGA 1572  
 Qy 528 AlalysTyrIleGluArgTyrAla 536  
 Db 1573 GCAAAATATGGAGAGGAAATTTGTGATAAATATTCTATATCATTTATCCTCCTGAG 1632  
 Qy 537 ---ArgLysLysHisAlaAspAsnAlaAlalysLeuHisSerLeuCys----- 551  
 Db 1633 CAGCAAAAAAAGTTTGTCTCTAAAAGTTCTGAAGAAAAAGAGCTGAGCATTTCTAAATTT 1692  
 Qy 552 -----GluAlaValLysThrArgAspIlePhe 560  
 Db 1693 GGGCAGGGGACCAAGTCAGAGCATCTCCCAAGTTTCAGTCAGAGTAATGAC---AGT 1749  
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
 Db 1750 GGAATTCAGCAGAGCTCTGTATGATGAGAGAGAACTTTTACCTCCACGGTGTGACCAAT 1809  
 Qy 581 Gly---HisGluProAsp----- 585  
 Db 1810 AGTTTATAGCCTGGAAGGAGAAAGCAAGATTCTTCTATGTTTCTGACTCGAAACAT 1869  
 Qy 585 ----- 585  
 Db 1870 CTTAATCAGACTTCAGCTTATATAGGCGCTCATATGAAAAAACCTTCCTAAATGGCT 1929  
 Qy 586 GluThrAlaLeuHisLeuAla----- 592  
 Db 1930 GAGGCTTTGGCTCATGGTGCAGAGCTGAACCTGGGCCAATTTCAGAGAAAAACAAGCGACA 1989  
 Qy 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610  
 Db 1990 CCACCTTATTCAGGCTGTATTAGGGGGCTCTTTGGTGACGTGTGAGTTCTCTCTACAGAA 2049  
 Qy 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630  
 Db 2050 GGTGTAATGTCACCAAGAGATGTCACAGGGCGGGACCAATTCACCATGCCACCGCTC 2109  
 Qy 631 ThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAla 650  
 Db 2110 TTAGGCGCACAGGGCAGGTATGTTTATTCCTAAAACGAGGTGCCCAATCAACATGCCACT 2169  
 Qy 651 AsnGluSerGlyLysThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670  
 Db 2170 GATGAAGAAGGAAAGACCCCTTTGACATAGCTGTGGAAGCAGCAGCAATGCTGATATAGTC 2229  
 Qy 671 GluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690  
 Db 2230 ACCTTGTACGTTTAGCA----- 2247  
 Qy 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710  
 Db 2248 -----AGAATGAATGAAGAGATCGGGAATCAGAAGGACTTTATGGA-----CAG 2292  
 Qy 711 ProSerGluAsnArgArgGluAspArgProLeuSerPheTyrGlnLeuGlySerAsn 729  
 Db 2293 CCAGGTGATGAACCTTATCAGGACATATTTCTGTGATTTTTCCTCCAAATGGCATCCAAT 2349

RESULT 13

US-10-080-334-79  
 ; Sequence 79, Application US/10080334  
 ; Publication No. US20040002584A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pena, Carol E. A.  
 ; APPLICANT: Shimkets, Richard A.  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Shenoy, Suresh G  
 ; APPLICANT: Kekuda, Ramesh  
 ; APPLICANT: Spytek, Kimberly A.  
 ; APPLICANT: Vernet, Corine A. M.  
 ; APPLICANT: Malyankar, Uriel M

Alignment Scores:  
 Pred. No.: 3,48e-33 Length: 2484  
 Score: 570.50 Matches: 212  
 Percent Similarity: 36.49% Conservatives: 143  
 Best Local Similarity: 21.79% Mismatches: 364  
 Query Match: 10.82% Indels: 254  
 DB: 17 Gaps: 24

US-09-914-042-1 (1-1006) x US-10-080-334-79 (1-2484)

Qy 23 AlaSerSerPheThrArgThrAlaGlnCys-----ArgAsn 35  
 Db 6 CCCGACGACATGACCGTGGAGTTCGAGAAAGTCGTCAGAACTCCCCCGCTTCAGGCGC 65

QY 36 ThrValAlaAlaIleGluGluAlaLeuAspValAepArgMetValLeuTyrIleMetLys 55  
Db 66 ACCATTGACGAGTGGAGCGACGCGTGTGGAGATTGAGCGCAAACTGGCAAGCTGGTG 125  
QY 56 LysSerValLysAlaIleAsnSerGlyLeuAlaHisValGluAsnGluGlnTyr 75  
Db 126 AAGCTGTGCGATGGTGAAGCCGGTAGGCTTACGTACGACACGACGAGGCTTTTC 185  
QY 76 ThrGlnAlaLeuGluLysPheGlyAsnCysValCysArgAspAspProAspLeuGly 95  
Db 186 GTGACGGCGTCCGCGACGCTGCCAGCAG-----TGCCAGGGCGACACGCTCATCTCG 239  
QY 96 SerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeu 115  
Db 240 GAATGCTCGCAGAGTTCCGTACAGCCTACAGGAGGTGGTGAACATACCATCATCTCCTG 299  
QY 116 IleGlnAsnMetAsnAsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeu 135  
Db 300 TTGTACACGAGCCGAGAGTCCGTGGCGAGCAGCTCCAGAGCTTTGTCAAGAGGATGTG 359  
QY 136 LysGlyValLysGlyAspLeuLysPheAspLysAlaTrpLysAspTyrGluThr 155  
Db 360 CGGAAGTTCAAG--GAGACAAAGAGCAGITTTGACAAGGTGCGGAGGACCTGGAGCTG 416  
QY 156 LysIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArg 175  
Db 417 TCCCTGGTG-----AGGAACGCCAGCGCCGAGGACCCGAGCAGCCGCCACGAG 461  
QY 176 ThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGluArgPhePheGln 195  
Db 462 GTGGAG-----GAAGCCACCGGGGCCCTCACCTCCACGAGAAAGTCTTCGCG 509  
QY 196 LeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIleLysGlyValAsp 215  
Db 510 CACCTGGCAGCTGACTATGTCTCCAGATCAATGTCTCGAGGCCAAGAGAGTTTGAG 569  
QY 216 LeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeu 235  
Db 570 ATCTGGAGCTATGCTGCTTCATGACGCGCCAGTCCAGCTTCTCCAGCGGCTAC 629  
QY 236 LysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIle 255  
Db 630 AGCTCTGTCACAGCTGACCCCTACATGAAGAAGCTGGCAGCGAGCTGGACACGAGTG 689  
QY 256 LysGlnAlaGlnAspGluArgGlnLeuIleGlnLeuArgAspIleLeuLysSer 275  
Db 690 GTGATCGACTGCGGTGAAAGCGTGAGTG----- 722  
QY 276 AlaLeuGlnValGluLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeu 295  
Db 723 -----GAGCGAAGCACGCGCCATCCAGCAGCGACCTTAGGACTTC 767  
QY 296 HisGlnProGlnGlyAsnLysGlu-----HisGlyThrGluArgAsn 309  
Db 768 TCCTACGATGATCCAAAGTGAGTTTGCCTGGACGCGCCCGAGTGGGTGGTGATGGAG 827  
QY 310 GlySerLeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSer 329  
Db 828 GCCTACCTCTTCAAGAGGCCAGCAACGCTTTCAGACATGAACCGGCGCTGGTTCTCC 887  
QY 330 ValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu 349  
Db 888 ATTCAGAACAGCCAGCTGTCTTACCAGAAGAAGCTCAAGGATGCCCTCACCGTGGTGTG 947  
QY 350 ---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGluLysLysCysPhe 366  
Db 948 GATGACCTCCGCTGCTGCTGTGAAGCCGTGTGAGGACATCGACGAGGAGTTCTGCTTC 1007  
QY 367 AspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGln 386  
Db 1008 GAGGTGCTGTACCCACCAAGAGCTGATGCTGCGAGGCTGACTCCGAGAAGCTGCGGCAA 1067  
QY 387 IleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu----- 400

Db 1068 GCCTGGTCCAGGCTGTGACAGCCAGCATCCGCTCCGCTACCCGAGAGCCCTGCACAGT 1127  
QY 401 -----AsnAsnAlaPheLysGlyAsp 407  
Db 1128 TGCTATAGCAGAGGCTGGACCGCACACGATCCCGCTCCAGCAGCAGCATCGACTCCGCC 1187  
QY 408 AspAsnThrGlyLysAsnAsnIleValGlnGluLeuThrLysGluIleSerGluVal 427  
Db 1188 ACCGACACTCGGAGCGTGGCGTGAAGGGCGAG-----AGTGTCTGCAGCGTGTG 1238  
QY 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447  
Db 1239 CAGAGTGGCGGCAACAGCAGCTGCGCGACCTGCGCGCAGCCCGCTGGGCC 1298  
QY 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467  
Db 1299 AGCATCAACCTGGGGGTGCTCTCTGCAATGAGTGCTCCGCAATCCAGAGGCTGGGT 1358  
QY 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487  
Db 1359 GTCCACTGCTCAAAGTGGCGTCCGCTGACGCTGAGACTCGTGGAGCCTGAGCTGCTAAAG 1418  
QY 488 LeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysLeuProAla 507  
Db 1419 CTGATGTGTGAGCTTGGAAACAGCGCTGTGAATCAGATCTATGAGGCCAGCTGTGAGGT 1478  
QY 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527  
Db 1479 CGAGCAGCAGCAAAACCCACAGCAGCAGCTCCCGGAGGACACAGGAGCCTGGATCAAG 1538  
QY 528 AlaLysTyrIleGluArgTyrAlaArgLys----- 538  
Db 1539 GACAAATACGTGGAAGAAAGAGTTCTCGGAAGGCGCCCATGGCACCAGCCCTGGAGGCC 1598  
QY 538 ----- 538  
Db 1599 CCAAGACGCTGGAGGGTGACAGAGTGCTGCGGCCCCACAGCTCTTCCCGCGCTCCCACT 1658  
QY 538 ----- 538  
Db 1659 GCCCGCCCAAGGTCCGGCTTGAGCCCGTCTGCGCTGTGTGGCGCTCTGTCTCTCAGAG 1718  
QY 539 -----LysHisAlaAspAsnAlaLysLeuHis 548  
Db 1719 GGTGACAGAGTCGAGGAGTCCAGCGGTGAGCAGAGCGGGACACTGAGCCGAGGCTGG 1778  
QY 549 SerLeuCysGlu-----AlaValLysThr 556  
Db 1779 GCGCTGCGCGACGTGCGGAGCTGCACCCGGGGCTTTGGCGCACCGCGCAGCGCTGCC 1838  
QY 557 ArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIle 576  
Db 1839 CGCGACCTTCTCGCTGCGCGCGCGCTGCGCCACCGGGCGAG-----GTC 1886  
QY 577 ProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVal 596  
Db 1887 AACTGGCGGAGCGGAGGATGAGGGCAAGACGCGCTG-----GTGACGCGCTG 1937  
QY 597 AspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLys 616  
Db 1938 CTAGGGGCTCTTGTATCGTCTGTGAGTTCTGTCTGCAAAACGAGGAGCGAGCTGAACCAA 1997  
QY 617 GlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCys 636  
Db 1998 AGAGACCGCGGGCGCGCGCGCGCTGACCGCAGCGCTGCTGGCGCGCGCGCGCGCAG 2057  
QY 637 LeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThr 656  
Db 2058 GTTTGCTCTTCTGAGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2117  
QY 657 ProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAla 676  
Db 676 -----

```
Db 2118 CCGTTGGCCATCGCA-----GTGACGGCG 2141
Qy 677 LeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGlu 696
Db 2142 CCCAAC----- 2147
Qy 697 AspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArg 716
Db 2147 ----- 2147
Qy 717 GluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaVal 736
Db 2148 -----GCTGACATCGTG 2159
Qy 737 SerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSer 756
Db 2160 ACACGTGCTCCGT--CTGCGCGCATGCGGAGAAATGCGCGAG----- 2201
Qy 757 IleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnPro 776
Db 2202 -----GCCAGGCTGCCCTCGTCTCCCGGGCGCCCTGGCG 2237
Qy 777 -AlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValG1 796
Db 2238 GGCAGCCCGCCAGAGCTCCAGTTCC-----GCAGGTGTATC 2273
Qy 796 nThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlyse 816
Db 2274 CAGGAGTTTCATCAGCTCCACC--TGGAAAG----- 2301
Qy 816 rArgGlnArgSerSerAspProProAlaValHisProProLeuProProLeuArgVa 836
Db 2302 -----AGAGTAGGCGGGCGGCGGCGGAGCTGCACCCCGCGCGCCGACGC-- 2352
Qy 836 lThrSerThrAsnProLeuThrProThrProProProValAlaLysThrProSerVa 856
Db 2353 -----CCGCGATGCCCGGAGTCCCTGGCGGCCCGCGCGCGC 2390
Qy 856 lMetGluAlaLeuSerGlnProSerLysProAlaProProGlylIleSerGlnIleArgPr 876
Db 2391 GGCCTGCTGTGACCGCGGTGCTATCTGCGAGCCCG-----AGTGTGGGGCGCC 2444
Qy 876 oProProLeuProProGlnProProSerArgLeuPro 888
Db 2445 CGGGCCCTGCTGCCCGCAGGAGAGCGAGGCCCGCCCA 2481

RESULT 14
US-10-176-306-10
; Sequence 10, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 10448-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
```

```
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (69)...(2570)
US-10-176-306-10

Alignment Scores:
Pred. No.: 2,07e-32 Length: 2737
Score: 561.00 Matches: 232
Percent Similarity: 37.13% Conservative: 136
Best Local Similarity: 23.41% Mismatches: 425
Query Match: 10.64% Indels: 199
DB: 15 Gaps: 32

US-09-914-042-1 (1-1006) x US-10-176-306-10 (1-2737)
Qy 20 AlaProThrAlaSerSerPheThrThrArgThrAlaGlnCys----- 33
Db 51 GCGCCGGCGCGCGCGCATGACCGTGGAGTTCAGAGAGTGGTCAAGGACTCCCGCGC 110
Qy 34 ---ArgAsnThrValAlaAlaIleGluGluAlaLeuAspValAspArgMetValLeuTyr 52
Db 111 TTCAGGCGCGCATTTGACGAGGTGGAGACGCGTGGAGATTGAGGCCAAACTGGAC 170
Qy 53 LysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGlu 72
Db 171 AAGCTGTGAAGCTGTGCAGTGGCATGTGTGGAAGCCGCTAAGGCTACGTGAGCCAGC 230
Qy 73 GluGlnTyrThrGlnAlaLeuGluLysPheGlyAsnCysValCysArgAspPro 92
Db 231 AGGCTTTTCGTGACCGCGGTCCCGACCTGTCACGAG-----TGCAGGGCGACACC 284
Qy 93 AspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeuPhe 112
Db 285 GTCATCTCGGAATGCTGTCAGAGGTTCGCTGACAGCGCTACAGGAGGTGGTGAATACCAC 344
Qy 113 LysAsnLeulleGlnAsnMetAsnIleIleSerPheProLeuAspSerLeuLys 132
Db 345 ATGATCTCTGTTGACCGAGGTCGCGTGGCGGAGAGCTCCAGAGCTTTGTCAAA 404
Qy 133 GlyAspLeuLysGlyValLysGlyAspLeuLysProPheAspLysAlaTrpLysAsp 152
Db 405 GAGGATGTGCGAAGCTTCAG--GAGACAAAGAGGAGTTTTCAGAGTTCGGGAGGAC 461
Qy 153 TyrGluThrLysIleThrLysIleGluLysLysGluHisAlaLysLeuHisGly 172
Db 462 CTGAGGCTGTCCCTGGTG-----AGGAACGCCCGAGGCCCGCCGAGCCGCG 506
```



```
Db 2420 CGTGACACTGCTCCGCTCTGGCGCGATGCGGAGGAATGCGGAGGCGGAGGCTGCCCC 2479
Qy 831 LeuProLeuArgValThrSerThrAsnProLeuThrProThrProProProVal 850
Db 2480 TGGTCCCGCGGCGC-----CTGGCGGCGAGCCCGCCACCGA 2515
Qy 851 AlalysThrProSerValMetGluAlaLeuSerGlnProSerLys----- 865
Db 2516 GCTCAGTTCGCGAGGTATCCAGGAGTTATCAGCCTCCACCTGGAGAGAGCTAGGG 2575
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProProLeuProProGlnProProSer 885
Db 2576 CCGGCGGCGCGGCGAGCTGCCACCCCGCGCGCCGAGCGCCCGCCGAGTCC 2635
Qy 886 ArgLeuProGlnLysLysProAlaPro-----GlyThrAspLysSerThr 900
Db 2636 CTGGCGCC--ACCGCGCGCGCGCTCGCTGTGACCGCGGCTCGATACCTGCGCAGCC 2692
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyPro 911
Db 2693 CCAGTGTGGGCGCGCGCGCTCTGCGCCCA 2725

RESULT 15
US-10-723-860-8033
; Sequence 8033, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8033
; LENGTH: 5421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1861)..(1865)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2487)..(2509)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-8033

Alignment Scores:
Pred. No.: 8 3e-32 Length: 5421
Score: 558.00 Matches: 256
Percent Similarity: 34.16% Conservative: 156
Best Local Similarity: 21.23% Mismatches: 422
Query Match: 10.58% Indels: 374
DB: 20 Gaps: 38

US-09-914-042-1 (1-1006) x US-10-723-860-8033 (1-5421)
Qy 12 AlaGluThrHisGluAspTyrLysAlaProThrAlaSerSerPheThrArgThrAla 31
Db 1743 GCTGAGGCTGCAGGTCACCCAGCCAGCC--CGTTCTGCTTTAGGCGCGACCATTCAC 1799
Qy 32 GlnCysArgAsnThrValAlaAlaIleGluAlaLeuAspValAspArgMetValLeu 51
Db 1800 GAGGTGGAGACGCGAGTGGTGGAGATTGAGGCCAAACTGGAC----- 1841
Qy 52 TyrLysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsn 71
```

```
Db 1842 ---AAGCTGGTGAAGCTGTCANNNNNATGGTGAAGCCGGTAAGGCTACGTCAGCACC 1898
Qy 72 GluGluGlnTyrThrGlnAlaLeuGluLysPheGlyGlyAsnCysValCysArgAspAsp 91
Db 1899 AGCAGGCTTTTCGTGAGCGCGCTCCGCGACCTGTCACAGAG-----TGCAGGCGGAC 1952
Qy 92 ProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeu 111
Db 1953 ACCGTATCTCGGAATGCTGACAGAGTTCGCTGACAGCCTACAGGAGGTGGTGAACCTAC 2012
Qy 112 PheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSerLeuLeu 131
Db 2013 CACATGATCTCTGTTGACAGGCCACAGGTCCTGCGGCGAGCAGCTCCAGAGCTTGTGTC 2072
Qy 132 LysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPheAspLysAlaTrrLys 151
Db 2073 AAAGAGGATGTCGGAAGTTCAAG--GAGACAAGAAGCAGATTGTGACAAGGTGCGGAG 2129
Qy 152 AspTyrGluThrLysIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHis 171
Db 2130 GACCTGGAGCTGCTCCCTGGTG-----AGGAACGCCAGGCCCGCCAGGCGAC 2174
Qy 172 GlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLysGluArg 191
Db 2175 CGGCCCGCCAGGCTGGAG-----GAAGCCACCGGGGCCCTCACCTCCACGAG 2222
Qy 192 ArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluLysIleLys 211
Db 2223 AAGTCTTCGCCACCTCGGACATGATGTCGACATCATGTTCTGCGAGCCCAAG 2282
Qy 212 LysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhePhe 231
Db 2283 AAGAAGTTGAGATCTCGGACTCTATGCTGCTTCATGCACGCCCGCTCCAGCTTCTTC 2342
Qy 232 GlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAsp 251
Db 2343 CAGCAGGCTACAGCTCTCTGCGACAGCTGGACCCCTACATGAAGAAGCTGGCAGCCGAG 2402
Qy 252 Leu-----HisThrIleLysGlnAlaGlnAspGluGluArgArgGln 265
Db 2403 CTGGACACGCTGGTGTGATCGACTCTGCGGTGGAAGCGTGAGATGGAGCGAAGCAGCC 2462
Qy 266 LeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSe 285
Db 2463 GCCATCCAGCAGCGGACGCTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2522
Qy 285 rGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGln 305
Db 2523 TGACGTG-----GACGCGCCAGTGG 2543
Qy 305 YThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrrGlnLys 325
Db 2544 GGTGCTGATGAGGCGCTACCTCTTCAAGAGGCGCCAGCAACGCTTTCAAGACATGGAACCC 2603
Qy 325 sArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgPr 345
Db 2604 GCGCTGTTCTCCATTGAGAACAGCCAGCTGGTCTACAGAGAAGCTCAAGGATGGCCT 2663
Qy 345 oProAlaLysLeu--AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGln 362
Db 2664 CACCGTGGTGGATGATCCCTCCGCTGCTGCTGAAGCCGCTGAGGAGACATCGAGCG 2723
Qy 362 uLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGln 382
Db 2724 GAGGTTCTGCTTCAGGTGCTGTCCACACAGAGCTGATGCTGCGAGGCTGACTCCGA 2783
Qy 382 uGlnGluCysGlnIleTrrMetSerValLeuGlnAsnSerLysGluGluAlaLeu----- 400
Db 2784 GAAGCTGGCAGACCTGGGTCCAGGCTGTGACAGGCCAGCATCGCTCCGCTCCCGGGA 2843
Qy 401 -----AsnAsnAla 403
```

2844 GAGCCTGACAGTTGCTATAGCGAGAGGCTGACCGCACAGCATCCCGTCCACGAGCAG 2903  
403 aPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluI 423  
2904 CATGACTCCGCGCACCGACACTCGGAGCGGTGGCGTGAAGGCGAG-----AGTGT 2954  
423 eIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAs 443  
2955 GCTGAGCGGTGTCAGAGTGTGGCGGCAACAGCAGTGGCGCGACTGCGCCACCGCGA 3014  
443 pProThrTirPLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHi 463  
3015 CCCCCTGGGCGACGATCACTCGGCGGTGCTCTGCTGATGAGTGTCTCGGATCA 3074  
463 aArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyTh 483  
3075 CAGGAGCCTGGGTGCTCCACTGCTCCAGAGTGGGTCCCTGACGCTGGACTCGTGGAGCC 3134  
483 rSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCy 503  
3135 TGAGCTGCTAAAGCTGATGTGTGAGCTTGGAAACAGCGTGTGAATCAGATCTATAGGC 3194  
503 sCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLys 523  
3195 CCAGTGTGAGGTGCGAGCAGCAGGAAACCCACAGCCAGCAGTCCCGCGAGGACAAGGA 3254  
523 sAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLys----- 538  
3255 GGCCTGGATCAAGGACAAATAGTGGAAAGAGTTTCTGGGAAGGCGCCCATGCGACC 3314  
538 ----- 538  
3315 AGCCCTGGAGGCCCCAAGACGCTGGAGGTGAGAAAGTGCCTGCGGCCCCACAGCTCTCC 3374  
538 ----- 538  
3375 CCGCGCTCCACTGCTCCCGCGCAAGTCCGGCTTGAGCCGCTTTCGCCCTGTGTGGCGCG 3434  
538 ----- 538  
3435 TCTGTCTCAGTGGGCACCTGGATCGTAAGTTCCGCGCGAGNACTCCCTCTTCTGTCCCGA 3494  
538 ----- 538  
3495 CGAGCTGGACTGCTTCTCTACTTCCAGCGAGGGCGCGAGGGGTGGCCCTCGCAA 3554  
539 -----LysHisAlaAspAsnAlaAlaLysLeuHi 548  
3555 GGGTGCAGAGTGGAGAGTCCAGCGGTGAGCGAGCGGGGACACTGAGGCGCGAGCGCTG 3614  
548 sSerLeuCysGlu-----AlaValLysTh 556  
3615 GGGCTCGGAGAGTGGCGAGTCCACCGGGGGCTCTTGGGGCACCGCGCAGCGCGTGC 3674  
556 rArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysI 576  
3675 CCGCGACCTTCTCGCTGGCGGGCGCTGCGCCCAACCGGGCGAG-----GT 3722  
576 aProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVa 596  
3723 CAACCTGGCGGAGCGGAGGATGAGGCAAGACGCGCTG-----GTGACGCGCT 3773  
596 lAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLys 616  
3774 GCTAGGGGCTCTCTTGATCGTCTGTGAGTTCTCTGCAAAACGAGCGGACGTAACCA 3833  
616 sGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCy 636  
3834 AAGACACGCGGGCGCGGCGCCCTGACACACCGCCAGCTGTGTGGCGCGCACCGGCA 3893  
636 sLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyLysTh 656  
3894 GCTTGTCTTCTTGAAGCGGGCGCGGACCCAGCACGCTGTGGACCAAGAGCAGCGGGA 3953

656 rProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAl 676  
3954 CCGTGTGCCATCGCA-----GTGACGCG 3977  
676 aLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGl 696  
3978 GGCAAC----- 3984  
696 uAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgAr 716  
3984 ----- 3984  
716 gGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaVa 736  
3985 -----GCTGACATCGT 3995  
736 lSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSe 756  
3996 GACACTGCTCGT---CTGGCGCGCATGGCGGAGAAATGCGCGAG----- 4038  
756 rIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnPr 776  
4039 -----GCCGAGGCTGCCCTCGTCCCGCGGGCGCCCTGCG 4073  
776 o-AlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValG 796  
4074 GGGCAGCCCCACGAGCTCCAGTTCC-----GCAGGTGTAT 4109  
796 lNThrAlaSerSerAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGly---- 814  
4110 CCAGGATTATCATGACCTCCACC---TGGAAAGAGAGCTAGGGCGGGCGAGCGCGGAGC 4166  
815 -----GlySerArgGlnArgSerSerSerSerSerProAlaValHisProL 831  
4167 TGCACCCCG 4223  
831 euProProLeuArgValThr-----SerThrA 840  
4224 CCGCGCGCTCGGTGTGACCGCGGTGATACCTGGCAGCGCCAGTGTGGGGCGCGCG 4283  
840 snProLeuThrPro----- 844  
4284 GCGCTGTGCTCGCGCGAGGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4343  
845 -----ThrProProPro-----ProValA 851  
4344 TTCCCTGGGGGGGTGCTGCTGCTGCGGTGCTCCCTCACCCCGCGCGCGCGCGCGCG 4403  
851 laLysThrPro-----SerValMetGluAlaLeuSerGlnProS 864  
4404 TCAGGTCACTCCCTTTTCAGGGCGCTGGGTGCTCATGTACAGCCACTTTTAAAGCGCG 4463  
864 erLysPro-----AlaProProGlyIleSerGlnIleArgPro----- 876  
4464 TGTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4523  
877 -----ProProLeu-ProProGlnProProSerArgLeuPro 888  
4524 AACACTGACCTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4580  
889 GlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnPro 908  
4581 GAAAGGTGGCGCATCCAGGGAACCAAGAGTACTCTACTGATGAC----- 4628  
909 ArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetVal 928  
4629 -----CCAGCATCCAGCTGCGGTGCTGGCGGGAGGGGCTCTGCGCTGGGT 4682  
929 LeuGlnPro-----ProAlaProMetProArg---LysSerGlnAlaThrLysLeuLys 945  
4683 CCTCAGCGCTGCTGCG 4741

Qy	946	ProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPhe	965
Db	4742	-----GTGCCAGACAGCCCGAGCCGATCCCGGCC	4771
Qy	966	SerGluGlyAspValIleIleValAspGlyGluGlnGluTyrTrpIleGlyHis	985
Db	4772	TCCTCCGAGGACCTTC-----TCCTGGTA-CTCGGCCCT	4806
Qy	986	IleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheValHis-----	1002
Db	4807	GTG-----CCGCGCATAAAGCCCTTTGCCACCTTGCAGCCCTCATCTCATGTTC	4857
Qy	1003	PheIleAlaAsp	1006
Db	4858	TTGTAGCGGAC	4869

Search completed: August 4, 2005, 17:14:51  
 Job time : 2828 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:33:01 ; Search time 393 Seconds  
(without alignments)  
4188.535 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAPVSVFVFIAD 1006

Scoring table:

	BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5		
Fgapop 6.0 , Fgapext 7.0		
Delop 6.0 , Delext 7.0		

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool h/US09914042/runat\_04082005\_090633\_2136/app\_query.fasta\_1.1159  
-DB=Issued Patents NA -QWTF=fastap -SUFFIX=rni -MINWATCH=0.1 -LOOPCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042@cgn 1 1 80 @runat\_04082005\_090633\_2136 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3874	73.5	2949	4	US-09-023-905A-8
2	3874	73.5	5954	4	US-09-023-905A-6
3	3691	70.0	4486	1	US-08-322-742-15
4	3417	64.8	4328	1	US-08-322-742-12
5	3155	59.8	2712	4	US-09-023-905A-11
6	3155	59.8	4595	4	US-09-023-905A-9
7	3013.5	57.1	5330	4	US-09-023-905A-1
8	2954.5	56.0	3456	4	US-09-023-905A-5
9	2954.5	56.0	4382	4	US-09-023-905A-3
10	571.5	10.8	6990	4	US-09-620-312D-155
11	453	8.6	1090	4	US-09-270-767-1629
12	453	8.6	1090	4	US-09-270-767-16911

13	429	8.1	2041	4	US-09-828-303-11	Sequence 11, Appl
14	347.5	6.6	712	4	US-09-270-767-11158	Sequence 11158, A
15	320	6.1	2119	4	US-09-620-312D-35	Sequence 35, Appl
16	291.5	5.5	7350	4	US-09-949-016-364	Sequence 364, Appl
17	284.5	5.4	3396	3	US-09-668-680-6	Sequence 6, Appl
18	280	5.3	3423	3	US-09-668-680-7	Sequence 7, Appl
19	279	5.3	213	4	US-09-513-999C-22310	Sequence 22310, A
20	268	5.1	6555	4	US-09-949-016-5295	Sequence 5295, Ap
21	261.5	5.0	2742	4	US-09-248-796A-1631	Sequence 1631, Ap
22	229.5	4.4	502	4	US-09-828-303-3	Sequence 3, Appl
23	229.5	4.4	2885	4	US-09-774-528-3	Sequence 3, Appl
24	223	4.2	1257	4	US-09-270-767-10482	Sequence 10482, A
25	221.5	4.2	5883	4	US-09-949-016-5001	Sequence 5001, Ap
26	221	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appl
27	219.5	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appl
28	217.5	4.1	1976	4	US-09-620-312D-515	Sequence 515, Appl
29	217	4.1	4378	2	US-09-080-897-3	Sequence 3, Appl
30	217	4.1	4378	3	US-09-323-735-3	Sequence 3, Appl
31	217	4.1	4399	3	US-08-899-595-2	Sequence 2, Appl
32	214.5	4.1	1978	4	US-09-949-016-5411	Sequence 5411, Ap
33	210.5	4.0	4403765	3	US-09-103-840A-2	Sequence 2, Appl
34	210.5	4.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
35	208.5	4.0	6202	4	US-09-774-528-120	Sequence 120, App
36	208	3.9	1698	4	US-09-599-287A-1	Sequence 1, Appl
37	206.5	3.9	39154	4	US-09-949-016-12384	Sequence 12384, A
38	206.5	3.9	39154	4	US-09-949-016-12801	Sequence 12801, A
39	206.5	3.9	39443	4	US-09-949-016-14326	Sequence 14326, A
40	206.5	3.9	39443	4	US-09-949-016-14327	Sequence 14327, A
41	206	3.9	1447	3	US-09-484-970B-121	Sequence 121, App
42	205.5	3.9	514	4	US-09-854-133-660	Sequence 660, App
43	205.5	3.9	229354	4	US-09-705-400-64	Sequence 64, Appl
44	204.5	3.9	8091	4	US-09-230-652-1	Sequence 1, Appl
45	204.5	3.9	8257	3	US-09-484-970B-65	Sequence 65, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-023-905A-8  
; Sequence 8, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Edding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; CURRENT FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 2949  
; TYPE: DNA  
; ORGANISM: Danio rerio  
US-09-023-905A-8

Alignment Scores:  
Pred. No.: 2.87e-249 Length: 2949  
Score: 3874.00 Matches: 755  
Percent Similarity: 83.12% Conservative: 92  
Best Local Similarity: 74.09% Mismatches: 122  
Query Match: 73.47% Indels: 50  
DB: 4 Gaps: 13

US-09-914-042-1 (1-1006) x US-09-023-905A-8 (1-2949)

Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
Db 1 ATGCTGACCAAGATAACAGTGGCGAGTTTGTACGGAGACAAATGAAGATTAAATCG 60  
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
Db 61 CCCACCGCCTCAAACTTCACCAACAGAAATGACTCACTGCAGGAAACACAGTATCCGCACTG 120  
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
Db 121 GAGGAGGCCCTGGATGTGGACCGCAGTGTCTTTTACAAGATCAAGATCAGTTAAGGCT 180  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
Db 181 ATTTACGCCCTCGGGTCTCGCTCATGTGGAGAAATGAGGACGAGTACACTCAAGCTCTGGAG 240  
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
Db 241 AAGTTCCGAGAGAACTGTGTGTACAGAGATGACCCGGACCTGGGATCAGCCTTCTCTGAAG 300  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db 301 TTCTCCGCTCTTCCACCAAGAGCTCACGCACTCTTCAAGAACTGTCTTCAAGAACATGAAT 360  
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 361 AATATCATTAACCTTCCCAATGGACAGTCTGCTGAAGGGAGAGTCTGAAAGGGGTAAAGGG 420  
Qy 141 AspLeuLysLysPropheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 421 GATCTCAAGAAAGCCCTTCGATAAAGCTGGAAGAGTACGAGAGCTAAAGTCTCTAAATA 480  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db 481 GAGAAGGAGAAAAAGAGCAGCCCGCAGCAGCAGGAATGATCCGAGCGAGATCAGCGGA 540  
Qy 181 AlaGluIleAlaGluLysGluLysGluLysGluArgPhePheGlnLeuGlnMetCysGlu 200  
Db 541 GCAGAGATAGCAGAAAGAGATGGAANAAGCGCGGTTCCTCCAGCTTCAGATGTGTGAG 600  
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 601 TACCTCTCTCAAAAGTCAATGAATCAAGATCAAAAAGGTGTGCGACCTCTCCAGAAATCTC 660  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 661 ATCAAAATCTTCCACGCACAGTGCAACTTCTTTTCAAGATGGTCTCAAAAGCGGTGACAA 720  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 721 CTCAAACCTCAATAGAAAAAATGCGCCACAGACTTGCACTCGATCAAAACAGGTACAGGAT 780  
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 781 GAAGAACCCACAGACAGTAAACCCAGTTACGGGATGTGCTAAAAACTGCTCTCAAGTGGAG 840  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 841 CAGAAGGAGGACTCTCAGGTAGACAGAGCGCCACCTACAGTCTGCAACCGCGCAGGCG 900  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 901 AACAAAGAGCATGGGACTTGAGCGCAGCGCAACCTTTTACAAGAAAGTGAAGCGGTGCGG 960  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 961 AAAGTGTGCAGAGAGAGAAAGTGCACAGTAAAGATGATATTTTGAACCATCTCACATGG 1020  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1021 ACGGCAAAACAGACCTCCCGCCAAACTCAATCTTCTCACCTGTGCAAGTGAAGCACAACCA 1080  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380

Db 1081 GAGGAGAGAAAAAGTTTTGACCTCATCTCATGACAGAACATATCATTTCCAGGCAGAA 1140  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1141 GATGAGCAGAGTGTCAAAATATGGATCTCAGTGTCTGCAGAACAGTAAAGAGAGCGGCTC 1200  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnIleValGlnGluLeuThr 420  
Db 1201 AACAAACCCCTTCAAGGCGCACAGCATGTTGGTGAATAATAACATTTGCGAGAGCTCACC 1260  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1261 AAGGCCATCTCGGAGAGGTGAAGCGGATGGCGGGAAACGATGTCTGCTGCGACTCGGT 1320  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1321 GCTCCGCGCCCAATGGCTCTCCCAACCTGGGCATCTCTGACCTGCATCGAGTGTTCG 1380  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1381 GGGATCCACAGAGAGCTGGGCGTCCATTACTTCCGAATCCAGTCCCTCACACTCGAGTC 1440  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1441 CTCAGCACTCCGAGCTTGTGTCGCCAAGAACGTGGGAATGTGTGCTTCAATGAGATC 1500  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1501 ATGGAGGCTGTCTGACGGCAGAAAGATGTGATCAACCGAATCCAGGCAGTGACATGCAG 1560  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaAaLysLysHis 540  
Db 1561 GCGAGGAAGGACTTTATATGTGCCAAATACACAGAAACGCTTCGCTCGTAAAGAGTGT 1620  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1621 CCAGACGCATGTGCGAAGCTGCACACGCTCTGTGATGCTGTGAAGGCCCGGACATTTTC 1680  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1681 TCTCTCATCCAGGTCTATGTGTGAAGAGTGTGATCTGATGGAGCCCATTTCTCTGGCTAAT 1740  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 1741 GGACATGAACAAGGTGAGACGGCTTTCATCTGCGCGTGAGACTGGTGGACAGAACTTCC 1800  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 1801 CTACACATCATCGACTTCTCCACCCAAAACAGTTTAAACCTGGATTAAGCAACCGCTAAA 1860  
Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLysLeuLeuLeu 640  
Db 1861 GGAAGCAGCTCTGTCATTTACTGTGCTGACGAGCAACAGCGAGTGTCTCAAACTGTGTG 1920  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 1921 CTCAGAGAAAGCCTCCATAGATATCGTAAATGAAGCTGGAGAGACCCCTTGGACATC 1980  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 1981 GCCAGGCGACTCAAAACATCTGCAGTGTGAGAACTGTCTGAACACAGGCTCTTTCAGGGAAG 2040  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2041 TTCAATGTCTATGTGCATGTGAGTATGAGTGGAGACTTCAGCATGAAGACTTCGACGAG 2100  
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2101 AGTGAAGATCTGGATGGAAGTCCAGTCTCT-----CACCGCGGATGAGCGGCCCC 2154  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740

Db 2155 ATCAGCTGCTACACACCGGAGTAACCTCCCTTCAGCTGAGTCCAGCAGCAGCTGAGCGGA 2214  
Qy 741 AspalalaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAen 760  
Db 2215 GACGCTGAGAGCTGGTTAAAGACAGCAAGCAAGC---TTTGTGCCAAACCTGGTCAACAAT 2271  
Qy 761 GluThrTyrglyAlaLeuLeuSerGlySerProProAlaGlnProAlaAaProSer 780  
Db 2272 GAAACCTACGGGACCATCATTAACACACAGCTCAACC-----GTGAGCGCTG 2316  
Qy 781 ThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSer 800  
Db 2317 TCCTCTCTGCTCCACTCTACACCCCGCAACCTA-----GTGAGCGCTGCTCTT 2370  
Qy 801 AlaAenThrLeu-----TrpLysThrAenSerValSerValAspGlyGly 815  
Db 2371 GCAGGACTGACTCAAGGATCTCCCGCTGGAAGCTGGCTCTCTGGATCTGAGCGGC--- 2427  
Qy 816 SerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArg 835  
Db 2428 ---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTGCGCTCTCTTACGG 2484  
Qy 836 ValThrSerThrAenProLeuThrProThr---ProProProValAlaLysThrPro 854  
Db 2485 GTCACTTCCACTCCCTCTTAATGCCAGCGGTGCTGCTCTCTGCTGCTAAAGCTACT 2544  
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyIleSer 872  
Db 2545 GGTATGAGGAGACCATGAATATGCACCCAAACCCGACAGGGGCTCTCTGGACAGAAC 2604  
Qy 873 GlnIleArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysPro 892  
Db 2605 ATCAACCGG----- 2613  
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 908  
Db 2614 GCTACAGTGGGACAAAACCTCAGCAAAAGCACACTGATGCGCTCC----- 2661  
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927  
Db 2662 ---GGATCCATCGAGACAGCAGCTAAGAGTCCAGGAGGCCCCACAAACACCACTGGT 2718  
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947  
Db 2719 CAAACTCTGCTGCGACCCACATGCTCCAGGAAA-----ACGTATTTGAAGCCGAAG 2769  
Qy 948 ArgValLysAlaLeuTyrglyAsnValAlaAspAsnProAspGluLeuThrPheSerGlu 967  
Db 2770 CGTGTGAAGGCCATGTATACTGTGTGCCGATATATCCAGACGAGCTGACCTTCTGTAG 2829  
Qy 968 GlyAspValIleIleValAspGlyGluGluAspGlnGluTrpTptIleGlyHisIleAsp 987  
Db 2830 GGAGAGCTTATCGTGTGATGGAGAGAGGAGGAGGAGTGTGGTGGGCCACATTTGAG 2889  
Qy 988 GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006  
Db 2890 GGAGAGCCAATGAGAGAGGAGCGTTCTCTGCTGACGTTGTACAGTTTCATTATGAC 2946

## RESULT 2

US-09-023-905A-6  
; Sequence 6, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; TITLE OF INVENTION: Therefor  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A

; CURRENT FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 5954  
; TYPE: DNA  
; ORGANISM: Danio rerio  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (433)....(3378)  
US-09-023-905A-6  
Alignment Scores: 7.5e-249 Length: 5954  
Pred. No.: 3874.00 Matches: 755  
Score: 83.12% Conservative: 92  
Percent Similarity: 74.09% Mismatch: 122  
Best Local Similarity: 73.47% Indels: 50  
Query Match: 4 Gaps: 13  
DB: 13  
US-09-914-042-1 (1-1006) x US-09-023-905A-6 (1-5954)  
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrllysAla 20  
Db 433 ATGCCTCACCAGATAACAGTGGCGGAGTTTGTCCAGGAGACAAATGAAGATTATAAATCG 492  
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40  
Db 493 CCCACCCTCAAACTTCACCACCAAGATGACTCCTCCTGAGAACACAGTATCCGACTG 552  
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrllysMetLysLysSerVallysAla 60  
Db 553 GAGGAGGCGCTGGATGGACCGCAGTCTCTTTACAAGATGAAGAGTCACTAGTTAAGSCT 612  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAenGluGluGlnTyThrGlnAlaLeuGlu 80  
Db 613 ATTTACGCTCGGCTCTGGCTCATGTGAGAAATGAGGAGCAGTACACTCAAGCTCTGGAG 672  
Qy 81 LysPheGlyGlyAenCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
Db 673 AAGTTCGAGAGAACTGTGTGTACAGAGATGACCCGACCTGGGATCAGCTTCTCTGAAG 732  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuIleGlnAsnMetAsn 120  
Db 733 TTCTCCGCTCTCCACAGGAGCTCACGGCACTCTTCAAGAACCTGTTTCAGAACATGAAT 792  
Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 793 AATATCATTACCTTCCCATTTGGACAGTCTGCTGAAGGAGAGATCTGAAAGGGGTTAAAGGG 852  
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrglyThrLysIleThrLysIle 160  
Db 853 GATCTCAAGAGCGCTTCGATAAAGCCTTGAAGAGCTACGAGACTAAAGTCTCTAAATA 912  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db 913 GAGAGAGGAGAAAAGAGACAGCCCGCAGCAGCGAATGATCCGACGAGATCAGCGGA 972  
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
Db 973 GCAGAGATAGCAGAGAGATGGAAGAGCGGGCTTCTTCCAGCTTCAGATGTGTGAG 1032  
Qy 201 TyrLeuLysValAenGluIleLysIleLysLysGlyValAspLeuLeuGlnAenLeu 220  
Db 1033 TACCTCTCAAGTCAATGAATCAAGATCAAAAGAGGTGTGACCTGCTCCAGATCTC 1092  
Qy 221 IleLysTyrglyPheHisAlaGlnCysAenPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 1093 ATCAATACTTCCACGACAGTGCAGTCTTTTTCAGAGTGTCTCAAGCGGTGAGCAAC 1152  
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260

1153 Db CTCAAAACCTCAATAGAAAACCTGCCACAGACTTGCACTCGATCAACAGGTACAGGAT 1212  
261 Qy GluGluArgArgGlnLeuLeuGlnLeuLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
1213 Db GAAGAACGACAGACAGTAAACCAAGTTACGGGATGTGCTAAAAAAGTCTCTGCAAGTGGAG 1272  
281 Qy GlnLysGluAspSerGlnLeuArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
1273 Db CAGAAGAGGAGACTCTCAGGTTAGACAGAGCGCCACCTACAGTCTGCACAGCGCGAGGCG 1332  
301 Qy AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
1333 Db AACAAAGAGCATGGGACTGAGCGCAGCGCAACCTTTACAGAGAGTGCAGCGGCTGGCG 1392  
321 Qy LysValTrpGlnLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
1393 Db AAAGTGTGGCAGAGAGAGAGTGCACAGTAAAGATGGATATTTGACCATCTCACATGGG 1452  
341 Qy ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
1453 Db ACGGCAACACAGACCTCCCGCAAACTCAATCTTCTCACCTGTCAAGTGAAGCACACCCA 1512  
361 Qy GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
1513 Db GAGGAGAGAAAAGTTTGGACCTCATCTCACATGACAGACATATCATTTTCCAGGCAGAA 1572  
381 Qy AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
1573 Db GATGAGCCAGAGTGTCAAAATATGGATCTCAGTGTGCAGAAACAGTAAAGAGAGCGCGTC 1632  
401 Qy AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
1633 Db AACAAACGCTTCAAGGGGACAGCATGTTGTGTAATAATCAATTTGTCAGGAGCTCAC 1692  
421 Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGly 440  
1693 Db AAGGCATCTCTGGGAGAGTGAAGCGGATGGCGGGAACGATGCTGTGCGAGCTCGCGT 1752  
441 Qy AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
1753 Db GCTCCGCGCCACATGGCTCTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTCG 1812  
461 Qy GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
1813 Db GGGATCCACAGAGCTGGGCGTCATTTACTCCCGAATCCAGTCCCTCACACTGCAGCTC 1872  
481 Qy LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
1873 Db CTCAGCACTCCGAGCTCTTGTGCCCAGAAACGTGGGGAATGCTGGCTTCAATGAGATC 1932  
501 Qy MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
1933 Db ATGGAGGCTGTCTGACGCGCAAGATGTGATCAAAACCGAATCCAGCCAGTGCATGCAG 1992  
521 Qy AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHis 540  
1993 Db CGAGGAAGGAGCTTTATCATGCCAAATACACAGAGAAACGCTTCGCTCGTAAGAAGTGT 2052  
541 Qy AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
2053 Db CCAGAGCGCACTGTCGAAGCTGCACAGCTCTGTGTGATGCTGTGAGGCCCGCGGACATTTTC 2112  
561 Qy GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
2113 Db TCTCTCATCCAGGTCTATGCTGAAGAGTGGATCTGTGAGGCCCATTTCTCTGGCTAAT 2172  
581 Qy GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
2173 Db GGACATGAACAAAGGTGAGACGCGCTTTCATCTGGCCGTGAGACTGGTGGACAGAACTTCC 2232  
601 Qy LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620

2233 Db CTACACATCATCGACTTCTCTCACCCAAAACAGTTTAAACCTGGATAAGCAAAACGGCTAAA 2292  
621 Qy GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
2293 Db GGAAGCAGAGCTCTGCAATTACTGCTGCTGACGACCAACAGCGAGTGTCTCAAACTGCTG 2352  
641 Qy LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660  
2353 Db CTCAGAGAAAGAGCTTCATAGATATCGCTNATGAGCTGGAGAGACCCCTTTGGACATC 2412  
661 Qy AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
2413 Db GCCAGGCGACTCAAAACATCTGCAGTGTGAGAACTGCTGAACACAGGCTCTTGCAGGGAAG 2472  
681 Qy PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
2473 Db TTCAATGCTCATGTGCAATGTGAGTATGAGTGGAGACTTCAGCATGAAGACCTCGAGCAG 2532  
701 Qy SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
2533 Db AGTGATGAGATCTGGATGAGAGTCAGTCTCT-----CACCGCGGGATGAGCGGCC 2586  
721 Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
2587 Db ATCAGCTGTACACACCGGCGAGTAACTCCCTTCAGCTGAGTCCAGCCAGCTGAGCCGA 2646  
741 Qy AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
2647 Db GACGCTCGAGACCTGGTTAAAGACACAGCAACGCG---TTTGTGCCAAACCTGGTCAACAAT 2703  
761 Qy GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780  
2704 Db GAAACCTACGGGACCATCATTAACACCGCTCACCC-----GTGAGCGCTG 2748  
781 Qy ThrThrSerAlaProProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800  
2749 Db TCCTCTTCTGCTCACCTCTACCAACCCCGAAACCTA-----GTTACGCGCTCTGCTCTT 2802  
801 Qy AlaAsnThrLeu-----TrrpLysThrAsnSerValSerValAspGlyGly 815  
2803 Db GCAGAGCTGACTCAAGGATCTCCCGCTGGAAGCGCTGCTCTCTGGATCTGAGCGGC--- 2859  
816 Qy SerArgGlnArgSerSerSerAspProAlaValHisProProLeuProLeuArg 835  
2860 Db ---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTGGCGCTCCCTTACGG 2916  
836 Qy ValThrSerThrAsnProLeuThrProThr---ProProProProValAlaLysThrPro 854  
2917 Db GTCACTTCCACTCCCTTCTAAATGCCAGCGGTGCTGCTCTCTCTCTGGCTAAAGCTACT 2976  
855 Qy SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyLysSer 872  
2977 Db GGATATGATGGAGACCATGAATATGCAACCCCAACCCGAGAGCGGCTCTCTGGACAGAAC 3036  
873 Qy GlnIleArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysPro 892  
3037 Db ATCAACCGG----- 3045  
893 Qy AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 908  
3046 Db GCTCAAGTGGCGGACAAAACCTTCAGCAAAAAGCACACTGATGCTGCTC----- 3093  
909 Qy ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927  
3094 Db ---GGATCCATCGAGACACAGCTAAAGAAAGTCCAGAGAGGCCCAACAAACACCATCTGGT 3150  
928 Qy ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947  
3151 Db CAAACTCTGCTGGACCCACATGCCCGAGAA-----ACGTATTTGAAGCCGAG 3201  
948 Qy ArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGlu 967  
3202 Db CGTGTGAAGGCCATGTATACTGTGTGCCGCGATATCCAGACGAGCTGACCTTCTCTGAG 3261



```
Db 1201 TGATGACGACATGATGAGAAATTCAGCCAGTCCCAACCGCGGGAGACCGGCCCAT 1260
Qy 721 eSerPheTyrGlnLeuGlySerAsnGlnSerAsnAlaValSerLeuAlaArgAs 741
Db 1261 CAGCTTTCACAGCTGGGCTCAACACAGCTTCAGTCAACGCTGTATCTTTGGCCAGAGA 1320
Qy 741 pAlaAlaSerLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnG 761
Db 1321 TGCTGCAAAACCTTGCCAAGGCAACAGAGGGCTTTTCATGCCAGCATCTTGCAGAAATGA 1380
Qy 761 uThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSerTh 781
Db 1381 GACTTACGAGCCCTTCCTGAGTGGAGCCACCTCCCGCCAGCTCGAGCCCGCCAGCAC 1440
Qy 781 rThSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSerA 801
Db 1441 CACGAGCGCCCGCCCGCTTCCTCCACGGAATGTTGGCAAAAGTTTCAGACAGCCTCCTCTG 1500
Qy 801 laAsnThrLeuTrpIlyThrAsnSerValSerValAspGlyGlySerArgGlnArgSerS 821
Db 1501 CTAACACCCCTGTGGAAAGCAAACTCTGTAAGTGTGGACGGTGAAGCCGGCAGCGATCTT 1560
Qy 821 eSerAspProProAlaValHisProProLeuProLeuArgValThrSerThrAsnP 841
Db 1561 CGTCAGATCCGCGAGCTGTCATCCACCGCTGCCCTCTTCGCGTGACATCTACCAATC 1620
Qy 841 roLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeuS 861
Db 1621 CCCTGACCCCAACGCGCGCCCGCCCGCTTCCACAGACGCCAGCGTAATGGAAGCCTTGA 1680
Qy 861 erGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProLeuProP 881
Db 1681 GCCAGCCGAGCAAGCCCTGCCCGCTGGGATCTCACAGATCAGGCCCGCCACCTCTGCCCC 1740
Qy 881 roGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrP 901
Db 1741 CACAGCCCGCCCGCCCTCCCGCAGAGAGAGCCCTGCCCGAGGGCTCACAAATCCACCC 1800
Qy 901 roLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuG 921
Db 1801 CACTGACCAACAAAGGCCAACCGAGAGACCTGTGGATCTCTCTGCAACGGGAAGCTCTGG 1860
Qy 921 lyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGlnA 941
Db 1861 GTCTCTCTGTCCAATGCTATGGTCTTCAGCCCGCCCTGCACCCCATGCTAGGAAGTCGCGAG 1920
Qy 941 laThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProA 961
Db 1921 CAACCAAGTTGAGCCTAAGCGGGTGAAGCGCTCTATAACTGTGTGGCTGACAAACCCCG 1980
Qy 961 spGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGluT 981
Db 1981 ATGAGCTCACCTTCTCCGAGGGGGATGTGATCATCGTGACGGGAGGAGGACAGGAGT 2040
Qy 981 roTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheV 1001
Db 2041 GGTGGATTGGCCACATTTATGAGATCTCTGGTTCGCAAAAGGCGCATTCCTCCGGTGTCA 2100
Qy 1001 alHisPheIleAlaAsp 1006
Db 2101 TGCATTTATCGCTGAC 2117
```

## RESULT 4

US-08-322-742-12

; Sequence 12, Application US/08322742

; Patent No. 5688641

; GENERAL INFORMATION:

; APPLICANT: Seger, Ruth

; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Fish &amp; Richardson

```
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,742
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/938,823
; FILING DATE: September 1, 1992
; APPLICATION NUMBER: 07/844,296
; FILING DATE: February 28, 1992
; APPLICATION NUMBER: 07/552,216
; FILING DATE: February 28, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/048003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-322-742-12
```

## Alignment Scores:

Pred. No.:	1,68-218	Length:	4328
Score:	3417.00	Matches:	648
Percent Similarity:	99.39%	Conservative:	1
Best Local Similarity:	99.23%	Mismatches:	3
Query Match:	64.80%	Indels:	1
DB:	1	Gaps:	0

US-09-914-042-1 (1-1006) x US-08-322-742-12 (1-4328)

Qy	355	GlnValLysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr	374
Db	3	CAGGTGAAGACCAACCCCTGAGGAGAGAAAGTGCTTTGACCTTATTTACATGACAGAACT	62
Qy	375	TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsn	394
Db	63	TACCACCTTCAAGCTGAAGATGAACAGGAATGTCAAAATATGGATGTCTGTCTGCAAAAT	122
Qy	395	SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsn	414
Db	123	AGCAAGAGAGAGCTTTAAACAATGCATTTAAGGGGATGACATCTCTGGAGAAAATAC	182
Qy	415	IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp	434
Db	183	ATCGTCCAAGAACTGACAAAGGAGATCATCTCAGAAAGTGACAGAGATGACGGGCAATGAC	242
Qy	435	ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu	454
Db	243	GTCTGTGTGACTGTGGGGGCCAGATCCTACATGGGCTTTCCACCAACCTGGGATCTCTG	302
Qy	455	ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln	474
Db	303	ACCTGCATCGAGTGTTCGGGAATCCACGAGAGCTGGGGGTTCATTATTCAGGATGACG	362
Qy	475	SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsn	494
Db	363	TCCCTGACCTTAGATGTACTGGGAACATCTCAGCTGTCTGCTCGCCAAAGATATTGGGAAT	422

QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsn 514  
Db GCAGGCTTAATGAGATCATGTAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAAC 482  
QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 534  
Db CAGGCGGAGCATGAATGCAAGAAGGACTATACATCAGCCAAAGTATCATGAGAGGAGA 542  
QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVal 554  
Db TACGCAAGAGAGAACACACCGGATAAACGGGGAAGCTTCACAGTCTTTGGAGGCCGTC 602  
QY 555 LysThrArgAspIlePheGlyLeuLeuGlnIleTyrAlaAspGlyValAspLeuThrGlu 574  
Db AAAACGAGAGATATTTTGGATGCTTCAAGCTTATGCTGATGTTGCTGATCTTACGGAA 662  
QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594  
Db AAAATCCCACTGGCCCAACGGACATGAGCGGATGAACGGCCCTCCACTTGCAGTCAGA 722  
QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614  
Db TCCGTGGATCGAACCTCTCTTCACATTTAGACTTTTATGTTTCAGAACAGTGGNACCTG 782  
QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634  
Db GATAAACAGACAGAGGAAAGGAGCAGACAGCCCTGCACTACTGCTGCTGACCGACAATGCC 842  
QY 635 GluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654  
Db GAGTGCCCTCAAGTTGCTCTCGGGGGGAGGCCCTCCATCGAGATAGCAAAATGAGTCAGGA 902  
QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674  
Db GAGACTCCGCTGGACATTTGCCAAGGCCCTCAAGACAGAGCACTGTGAGGAGCTGCTGACC 962  
QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeu 694  
Db CAAGCCTTATCTGGAAGATTTAAATCTCACGTTACGTTGAATGAATGATGCGACTACTC 1022  
QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714  
Db CACGAAGACCTGGATGAAAGTGATGACACATGGATGAGAAATTCAGCCCGCTCCCAAC 1082  
QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734  
Db CGGCGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCAACACGCTTCAGTCTAAC 1142  
QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754  
Db GCTGTATCTTTGGCCAGAGATGCTGCAAACTTGGCAAGGACACAGCAGAGGCTTTTCATG 1202  
QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774  
Db CCCAGCATCTGCAAGATGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCACCTCCCGCC 1262  
QY 775 GlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValGlyLys 794  
Db CAGCTTGAGCCGCCAGCACCCAGCCGCCCTCCCGCTTCTCCACGGAATGTGGCA 1322  
QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValAspGly 814  
Db AGTTTCAGACAGCCTCTCTGCTTAACACACCTGTGGAAGACAAACTCTGTAAGTGTGACGG 1382  
QY 814 yGlySerArgGlnArgSerSerAspProProAlaValHisProProLeuProProLe 834  
Db TGGAGCCGGCAGCGCATCTTCTGATCCGCGCAGCTGTCCATCCACCGCTGCCCGCTCT 1442  
QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPr 854  
Db TCGCGTGATCTTACCAATCCCTGACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1502

QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerGlnIle 874  
Db CAGCGTAATGGAAGCTTTGAGCCAGCCGAGCAAGCTGCCCCCGCTGGGATCTCACAGAT 1562  
QY 874 eArgProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894  
Db CAGGCCCCCCTCTGCCCCCAGCCGCGCCCTCCCGCAAGAGAGCTTGGCC 1622  
QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLe 914  
Db AGGGGCTGACAAGTCCACCCCTGACCAACAAAGGCCAACCGAGAGACCTGTGATCT 1682  
QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934  
Db CTCTGCAACGGAAGCTCTGGGTCTCTGTCCTGTCCTGTCGAGCCCTGTCACC 1742  
QY 934 oMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAs 954  
Db CATGCTTAGGAAGTCCGAGGCAACCAAGCTTAAGCGGTGAAGCGCTCTATAA 1802  
QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAs 974  
Db CTGTGTGCTGACAACCCCGATGAGTCACTTCTCCGAGGGGATGTGATCATCGTGA 1862  
QY 974 pGlyGluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysG 994  
Db CGGGGAGGAGGACGAGGAGTGTGGATTGGCCACATTTGATGAGATCTCTGTGCGAAAG 1922  
QY 994 yAlaPheProValSerPheValHisPheIleAlaAsp 1006  
Db CGCATTCGGGTGTCTTTGTGCACCTTTATCGCTGAC 1959

## RESULT 5

US-09-023-905A-11  
; Sequence 11, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; TITLE OF INVENTION: Therefor  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; CURRENT FILING DATE: 1998-02-13  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 2712  
; TYPE: DNA  
; ORGANISM: Danio rerio  
US-09-023-905A-11

Alignment Scores:  
Pred. No.: 2,65e-201 Length: 2712  
Score: 3155.00 Matches: 621  
Percent Similarity: 75.60% Conservative: 98  
Best Local Similarity: 65.30% Mismatches: 174  
Query Match: 59.83% Indels: 58  
DB: 4 Gaps: 10

US-09-914-042-1 (1-1006) x US-09-023-905A-11 (1-2712)

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
Db 1 ATGCTCATCAGCTGACAGTACTGAGTTGTGGATATTACCATGAGGACTATTAAGCA 60  
QY 21 ProThrAlaSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaIle 40



Db CCGCAACATCAGTGTCTGCGCGCATGGCTCACTCAGGAATACAGTCCGCGCTCTG 120  
Qy GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
Db GAAGAGCGCTGGATCTGCGCGCATGTCTGTCACAAATGAAGATCAGTCAAGGCC 180  
Qy IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80  
Db ATAAACAGCTCTGGTCAGACTCATGTAGAACAGGAGGAGCATCATCCAGGCCATAGAG 240  
Qy LysPheGlyGlyValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
Db AGTTTACGGATAACACTGTGTCAAAAGATGACCTTGAGATGTCCTCAATCTCTCTACA 300  
Qy PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
Db TTGCGTGGTTTCAACAAGAGCTTACTGCTCTTTTCAAGAACTTGCTACAGAACTGAAT 360  
Qy AsnIleIleSerPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle 160  
Db AACATCATCACTTTTCCACTAGACAGTCTGCTAAAGGGAGACCTCAAGGGAGTCAAGGG 420  
Qy AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle 160  
Db GATTTGAAAAGCCATTTGATAAAGCATGGAAGGATTAAGAAACCAACTGAGCAAGATT 480  
Qy GluLysGluLysLysGluHisAlaLysLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db GAGAAAGAAAAGCGAGAAACATGCCAAACAGACAGCGGTCTGATCCGAACAGAGATCAGTGA 540  
Qy AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
Db GGAGAGATCGCAGAGAGATGAGAAAGAGAGACCGCTGTTTCAGCTTCAGATGTGTGAG 600  
Qy TyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db TACCTATTAAAGTGAATGAATCAAGTCAAAAAGGGGGTGCACCTGCTTCACAACTC 660  
Qy IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db ATCAAAATCTTTTCATGCCAGTGCATTTCTTTTCAGGATGGCTTAAAGTCTGTGCAACT 720  
Qy LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db CTGAACCTTTTCATGAAAAGCTTGCCACAGACTTAACCGCGAAACAAACAGACTCAAGAT 780  
Qy GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db GCAGAAAGGAACAGTGTGTCAGCTGAAAGAACTCTTAAATCTGCTCTACAGTCTGAG 840  
Qy GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db TGTAAAGGAGGATGCTCAGTCAAGCAGACAGCAGCGCTACAGTCTTCACCGAGTTCGAGGC 900  
Qy AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db AATAAGCTCAGCGCAGCGAGCGCTCTGGGATGCTCTCTCAACCGCAGCGAGGAGCTGAGG 960  
Qy LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db AAAAGTTGGCAGAAAAGGAAGTCTGTGAAAATATGGATTGTTGACTATTTCACATGGA 1020  
Qy ThrAlaAsnArgProProAlaLysLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db ACGCCCAATGCACCGCCAGCAACCTGAACCTCTTAAACCTGCAAGTGAAGCGTAAACCCA 1080  
Qy GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db GATGAGAAAAAATGCTTTGATCTCATATCATATGACAGAACGTATCATCTTCCAGACTGAG 1140  
Qy AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400

Db GATGAGCGAGAGTGTCTCAGGTATGGGTTTCTGTCTCCAGAACAGTAAAGAGAGCGCGCTG 1200  
Qy AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db AACAAATGCTTTAAAGACGATCAGAATGAGGGAGAAAATAACATTTGTTTCGAGAGCTCACT 1260  
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db AAGGCCATCTGGGGGAAGTGAAGAAATGAGCGCAATGACGTGTGCTGTGCTGTGGA 1320  
Qy AlaProAspProThrTyrPheLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db GCTTCCAATCAACATGGCTCTCCACAAACCTGGGTGTGTTGATTGTCATTTGAATGCTCT 1380  
Qy GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db GGGATCCATCGGAAATGGGCGTCCACTACTCCGAATACAGTCTCTGACACTGGACCTC 1440  
Qy LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db TTAGGCACATCTGAATATTGCTTAACAGTGTGGGAATGAGCATTTCAATGAATC 1500  
Qy MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db ATGGAAGCAAACTGTCTTCAGAG---ATCCCAAAACCTACCTTCTAGTGACATGCAG 1557  
Qy AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540  
Db GTACGAAAGAACTTCACTACAGCCAAATACACAGAAAGCGTTTCGGCTCAGAAAGATAT 1617  
Qy AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db GCAGATTAAGCAGCTCGACTGCATGCATGCTGTGATGCAGTGAAGTCTCGGACATCTTC 1677  
Qy GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db TCCCTGATCCAGGCTATGCTGAAGGACTGGACCTGATGGAGACCAATTAATCAGCCTAAC 1737  
Qy GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db CAAATGAACCAGCGCAGACATCATCTATCTTGGGTACGAATGGTGGACCGAACTCC 1797  
Qy LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db CTCCATATTGTGGACTTCTTGACAGAACAGTGGCAATTTAGACAAGCAGACAGCCAAA 1857  
Qy GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db GGAAGCAGCGCTACATTTATTGCTGTGACTGATTAACAGTGAATGTTATGAAGCTGCTG 1917  
Qy LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db CTGGGGGGAAGGATCTGTGAGCATTAATGATCTGGAGAGACTGCTCTGGATTG 1977  
Qy AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db GCGCAGCGCTCTCAAACTCCAAATGCGAGGAGTGTGCTGACTCAGGCGCAGACGGGAAG 2037  
Qy PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700  
Db TTCAATGTCCATGTGCATGTGGAATATGACTGGGCTCTGTCATTAATGAGGATCTGGACGAG 2097  
Qy SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db AGCGAAGATGAGATGGAGGACAAG-----CCCATTCCTCCATCAGCGCTGAGAGCGTCCA 2151  
Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db ATAGCTATAGTTCAGGCGAGTGGCCCCCATGATGCCCAACATCAGCGCTCTGGCTCGG 2211  
Qy AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db GACGTGGCCAAATGTGGTCAATAAAGCAGAGGGGCTTTTATTCCGAGCATGATGAAC 2271





Db 1260 AAAGTTTGGCAGAAAAGGAAGTCTCTCTGTAATAATGGATTCTTGACTATTTCACATGGA 1319  
Qy ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
Db 1320 AGCCCAATGACCCGCCAGCAAACTGAACTCTTAACCTCGCAAGTGAAGCGTAACCCA 1379  
Qy GluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1380 GATGAGAAAAAATGCTTTGATCTCATATCATGACAGACGATATCACTTCCAGACTGAG 1439  
Qy AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGlnAlaLeu 400  
Db 1440 GATGAGGAGAGTGTGAGTGTGGTCTCTGTTCTTCCAGAAACAGTAAGAGAGCGCTG 1499  
Qy AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 1500 AACAAATGCTTTAAAGACGATCAGAAATGAGGAGAGAAATAACATTTGTCGAGAGCTCACT 1559  
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGly 440  
Db 1560 AAGCCATCGTGGGGAAGTGAAGAAATGAGCGCAATGACGTGTGTGACTGTGGA 1619  
Qy AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1620 GCTTCCAATCCAACATGGCTCTCCACAAACCTGGGTGTGTTGATTTGCATTGAATGCTCT 1679  
Qy GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1680 GGGATCCATCGGGAATGGCGTCCACTACTCCCGAATACAGTCTCTGACACTGGACCTC 1739  
Qy LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1740 TTAGGCACATCTGAATATTGCTTGAACAGTGTGGGAATGACGATTCATGAATC 1799  
Qy MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1800 ATGGAAGCAAACTCTCTTTCAGAG---ATCCCAAAACCTACCTCTCTAGTGACATGCAG 1856  
Qy AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysHis 540  
Db 1857 GTACGAAAGACTTTCATCACAGCAAAATACACAGAGAAGCGTTTCGCTCAGAGAAGTAT 1916  
Qy AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1917 GCAGATACCGCAGCTCGACTGCATGCTGTGTGATGAGTGAAGTCTCGGACATCTTC 1976  
Qy GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1977 TCCCTGATCCAGGTCTATGCTGAAGGACTGGACCTGTGAGACCATTAATCAGCCTAAC 2036  
Qy GlyHisGluProAspGluThrAlaLeuHisIleuAlaValArgSerValAspArgThrSer 600  
Db 2037 CAACATGAACCGGCGAGACATCATCTATCTTGGGTACGAATGGTGGACCGAAACTCC 2096  
Qy LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 2097 CTCCATATTGGGACTTCTTGATACAGAACAGTGGCAATTTAGACAGACACAGCCAAA 2156  
Qy GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 2157 GGAAGCAGCGCTACATATTGCTGCTTACTGATGATAACAGTGAATGATGAAGTGTG 2216  
Qy LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 2217 CTGGGGGGAAGCATCTGTGACGATTTACTAATGATGCTGGAGAGACTGCTGTGATTTG 2276  
Qy AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2277 GCGAGCTCTCAACACATCCAAATGCGAGGAGCTGCTGACTCAGGCCACAGCGGGAG 2336  
Qy PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
Db 2337 TTCAATGCCATGTGCATGTGGAATATGATGCGCTCTGCATTAATGAAGGATCTGGACGAG 2396

Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2397 AGCGAAGATGAGATGGAGACAAG-----CCCATTCCTCAGCGGTGAGGAGGTCCA 2450  
Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2451 ATAAGCTGTATAGTTTCAGGCGAGTGGCCCATGATGCCAACATGAGCGCTCTGGCTCG 2510  
Qy AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
Db 2511 GACCTGGCCAATGGTCAATAATGAACAGAGGCGCTTTATTCGAGCATGATGAAC 2570  
Qy GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780  
Db 2571 GAGACTTACGGCACCATGCTCGATCCCACTCTCCACACTG-----GGTTTACCAGGA 2624  
Qy ThrThrSerAlaProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800  
Db 2625 GTACCTGGCATTCCTCTTTTACCCCTCGGCCCTTGGAGGGGA----- 2669  
Qy AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820  
Db 2670 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGGTCA 2708  
Qy SerSerAspProProAlaValHisProProLeuProLeuArgValThrSerThrAsn 840  
Db 2709 TGTTCAATCTGCA-----AACCTCAAACTCTGNACAAAATAACTCTGTGTAT 2759  
Qy ProLeuThrProThrProProValAlaLysThrProSerValMetGluAlaLeu 860  
Db 2760 GTTCTGCTCTGCTCTCTCCACCTCTCTGCA----- 2792  
Qy SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880  
Db 2793 -----CCCAAGAGACCTCCACCTCCAGATCCAAAGGCCAGTCTTCTCTCCAGCAGCC 2846  
Qy ProGlnProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900  
Db 2847 ACGCTCT-----CTGCACCA-----TCCGCA 2870  
Qy ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920  
Db 2871 CCGCTCTTATTCACCTGCTCTCTCTCAGGCCAGCGCT----- 2909  
Qy GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940  
Db 2910 -----GTAGTGGCCCTCGACCCAGTTATGCCCACTTCGTCA 2945  
Qy AlaThrLysLeuLysProLysArgValLysAla 951  
Db 2946 CTGACTGATGTCAAAAGTCTGCTGTCTTAAGCC 2978

## RESULT 7

US-09-023-905A-1  
; Sequence 1, Application US/09023905A  
; Patent No. 6475778  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Thomas M.  
; APPLICANT: King, Frederick J.  
; APPLICANT: Harris, David F.  
; APPLICANT: Hu, Erding  
; APPLICANT: Spiegelman, Bruce  
; APPLICANT: Chan, Joanne  
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses  
; FILE REFERENCE: DFN-021  
; CURRENT APPLICATION NUMBER: US/09/023,905A  
; CURRENT FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/038,191  
; PRIOR FILING DATE: 1997-02-14  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1  
 ; LENGTH: 5330  
 ; TYPE: DNA  
 ; ORGANISM: Bos sp.  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (209)...(3595)  
 US-09-023-905A-1

## Alignment Scores:

Pred. No.: 1,88e-191 Length: 5330  
 Score: 2013.50 Matches: 627  
 Percent Similarity: 66.67% Conservative: 129  
 Best Local Similarity: 55.29% Mismatches: 225  
 Query Match: 57.15% Indels: 153  
 DB: 4 Gaps: 23

US-09-914-042-1 (1-1006) x US-09-023-905A-1 (1-5330)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
 Db 269 ATCGCGACACAGATCTCCGTCCTCCGAGTTTCATCGCGGAGACACCGAGGACTTACAACCTCG 328  
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
 Db 329 CCCACCACGTCACGCTTCACTACGCGGCTGCACAACTCGAGAAACACCGTCAACGCTGCTG 388  
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
 Db 389 GAGGAGGCTCTAGACCAAGATAGAACAGCCTTACAGAAAGTTAAGAACTCTGTAAAGCA 448  
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80  
 Db 449 ATATACAATTCGGTCAAGACCATGTACAAATGAAGAAACTATGCGCAAGTCTCTTGAT 508  
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
 Db 509 AAGTTGGGAGTAATTTTAAAGTCGACAGAACCCAGATCTTGGCAGCGCTTTTGTCAAG 568  
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 Db 569 TTTTTCAGCTTACAAAGAACTGTCCACACTGTCTGAATAATCTCTCCAGGCGCTGAGC 628  
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140  
 Db 629 CACAATGTGATCTTCACTTGGATTCCTTTGTGAAGGAGACCTTGAAGGAGTCAAGGC 688  
 Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
 Db 689 GATCTCAAGAAACCATTTGACAAAGCTTGGAAAGATTATGAGACCGAAGTTTACCAAAAT 748  
 Qy 161 GluLysGluLysLysGluHleAlaLysLysGlyMetIleAspThrGluIleSerGly 180  
 Db 749 GAGAAAGGAGAGAGGAGGACGCCAAGCAGCAGCGGATGATCCGACCGAGATCAACCGC 808  
 Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
 Db 809 GCCGAGATCGCGAGGAATGAAAGGAGGCGCGCTCTTCCAGCTCCAGATGTGCGAG 868  
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220  
 Db 869 TATCTCATTAAGTTAATGAATCAAGACCAAAAGGCTGTGATCTGTCGAGAACCTG 928  
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240  
 Db 929 ATAAAGTATATACGACAGTGAATTTCTTTCAAGATGGTTTGAAGACAGCTGATAA 988  
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 Db 989 TTGAACACAGTACATTGAAAGCTGCTGCTGATTTGTATATATATCAACAGACCCAGGAC 1048  
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280

Db 1049 GAAGAAAGAAACAGCTGACCGCACTCTCGAGACCTAATAAAGTCTCTCGCTCCAACCTCGAT 1108  
 Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
 Db 1109 CAGAAGGAGTCTAGGAGAGATTCCAGAGCGCGAGGGA---GGCTACAGCATGCACCAG 1165  
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
 Db 1166 CTGCAAGGCAACAAGGAATACGCGCAGAGAGAGGAGGCTACCTGCTGAAGAAGAGTGAC 1225  
 Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
 Db 1226 GGGATCCGGAAGGTGTGCAGAGAGGAGTCTCCGTCAGAACGGGATCTCTGACCATC 1285  
 Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
 Db 1286 TCCCAAGCCACGTCACACAGACAGCCAGCTGAACCTTCTCACITGCGAGGTGAAG 1345  
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377  
 Db 1346 CCGAATCGCGAGGACAGAGTCTTTTGACCTGATATACATACAGGACGTATCACTTT 1405  
 Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
 Db 1406 CAGCGCGAAGATGACGAGGATTATGACGTGGATCTCGGTGCTGACAAACAGAAAGAG 1465  
 Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417  
 Db 1466 GAGGCCCTCACCATGCGCTTCCGGGGGGAACAGAGTGTGGGGAGAGCAGCTG---GAG 1522  
 Qy 418 GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
 Db 1523 GAGCTGACGAAGGCCATCATCGAGGACGTGCGAGCGCTCCCGGGCAACGAGCTGCTGTC 1582  
 Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457  
 Db 1583 GACTGCGCTCGGAGAACCCACCTGCTGCTCCACCACTTGGGCATCTTGACCTGTATA 1642  
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGluSerLeuThr 477  
 Db 1643 GAATGTTCCGCGATCCATAGAGAAATGGGGTTCATATTTCTCGCATCCAGTCTTTGGA 1702  
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
 Db 1703 CTAGACAAATAGGAATCTTGAATCTTGTGCGCAAGAAATGTAGGAACAATAATGTTT 1762  
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
 Db 1763 AATGATATTATGGAAGCAAAATTTTACCCAGT---CCCTCACCACCAACCCACCCCTCAAG 1819  
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537  
 Db 1820 GATATGACTGTACGAGGAGATATATCTCTGCAAGATATGTAGATCATAGTTTCACGG 1879  
 Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
 Db 1880 AAGACCTGTCATCGTCATCAGTAACTGAACGAATTTGCTTGAGGCCATCAATCCAGG 1939  
 Qy 558 AspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
 Db 1940 GATTACTTGACATAATTTCAAGTCTATGCAAGGGGGTGGAGCTAATGGAACCCGCTG--- 1996  
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
 Db 1997 CTGGAAACCGGACAGGAGCTTGGGAGACAGCCCTTCTCATCTTGCATGCTCGAACCCGAC 2056  
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
 Db 2057 CAGACATCTCTCCATTTGGTGGCTTCTCTGTCACAAACTGTGGGAACCTAGATAAGCAG 2116  
 Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeu 637  
 Db 2117 ACGCCCTTGGGGAACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2176

```
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657
Db 2177 AAGTGTCTGCTCAGGAGCAGCCCACTGTGACGTCGTTAATCAGCTGGAGAGCCGCC 2236
Qy 658 LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677
Db 2237 CTGGACATAGCAAGAGAGACTGAAGCCCACTCAGTGTGAAGACCTGTCTTCCAGCTAAA 2296
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697
Db 2297 TCTGGAAGTTCATTCACACGTCCACGTCCAGTGAATATGATGGAATCTTTCGACGAGGAG 2356
Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGlu 717
Db 2357 ATGATGAGAGGATGACGACTGTGATGACAAACCGACCCCATCAAGAGGCGCTCC 2416
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737
Db 2417 CCCGACCGCAGAGCTTCTGCCACTCTCCAGCATCTCCCCCGAGGAC-----AAG 2467
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757
Db 2468 CTCTCACTGCCGGGCTTCAGCAGCGCAAGGACAAACGACTC----- 2512
Qy 758 LeuGlnAsnGluThrTyrGlyAla-----LeuLeuSerGlySerPro 771
Db 2513 -----TCTACGGCGCCTTCACCAACGACATCTCGTCTCCACAGCACAGAC 2560
Qy 772 ProProAlaGlnProAlaProSerThrThrSerAlaProProLeuProProArgAsn 791
Db 2561 TCACCCACGTCAACCGATCGCA-----GAGGCGCCCGCTGCTCCCGAGAAC 2608
Qy 792 ValGlyLys-----ValGlnThrAlaSer 799
Db 2609 GCCAGAAAGGTTCACCTGGCCACCTTCAACACTCTCTAAGCACCCAGACCTCTAGT 2668
Qy 800 SerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819
Db 2669 GGCAGCTCAACCTGCTCCAGNAGCGGTCTCTCCCCACACCACCGGA---CACAAAGA 2725
Qy 820 SerSerSerAspProAlaValHisProProLeuProProLeuArg-----ValThr 837
Db 2726 ACCCTGTCTGACCTCCCGCCCACTACTCTACGGGCCCCCAACAAAGCGCAGTTCTCT 2785
Qy 838 SerThrAsnProLeuThrProProProProProValAlaLysThrProSerValMet 857
Db 2786 TGGGTAACGAGCTGGGTCCCTCATCGTCC-----AGTAAGACCAACGAACAAAGTTC 2836
Qy 858 GluAlaLeuSerGln----- 862
Db 2837 GAGGCGCTTCCACAGCTCGNAGCACCAGTTCGCAAGACTGCACTTGTCCCAAGAGTT 2896
Qy 863 -----ProSerLysProAla----- 867
Db 2897 CTTCCTAAACTACCTCAGAAAGTGGCACTAAGAGAAACAGAGACCAAGCATCTCTCC 2956
Qy 868 -----ProProGlyLe-----SerGlnIle----- 874
Db 2957 CTCGACAAAGCAAGCTCCCACTGAGATCTTCCAGAAAGTGTCTCCAGTTGACAGAGTTA 3016
Qy 875 -----ArgProProPro-----LeuProProGlnPro----- 883
Db 3017 CGCGAGAAGCCCAACCGGGGACCTGCCCCGAAGCCCAACGGAAGTGGCTCCCAACCC 3076
Qy 884 -----ProSerArgLeuProGlnLys----- 890
Db 3077 CCCATTGGAGACTTACCACCTAAGCCAGCGAGCTGCCCCGGAAGCCACAGCTGGGCGAC 3136
Qy 890 ----- 890
Db 3137 CTGCCCCCAAGCCCGAGCTCGCAGACTTTGCCCCCAAGCCCGCAGGTGAAAGACCTGCT 3196
```

```
Qy 891 ---LysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909
Db 3197 CCCAAGCCACAACACTGGGGAGCTGCTGCAGAAAAACCCAGAGAGAGCGCTCCGCCAAG 3256
Qy 910 Gly-----ProValAspLeuSer----- 915
Db 3257 GCCCAGCCACCCCTGGAGCTCACCCCAAGTTCACACCGCGGAGACCTGTCTCCCGAACGTC 3316
Qy 916 -----AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProPro 932
Db 3317 CCCAAGCAGGCGTCTGAGGACCAACAGACCTCAGCCGCCACCTGCCA---GAGACACCC 3373
Qy 933 AlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeu 952
Db 3374 GTGCCTCTGCCAGAGATCAACACCGGGAAGAGCAAGGTGAGCGAGTGAAGACCATC 3433
Qy 953 TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIle 972
Db 3434 TACGACTGCCAGGCGCAACGATGACGAGCTGACTTTTCATGGAGGCGAGGTGATCGTG 3493
Qy 973 ValAspGlyGluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyVal 992
Db 3494 GTCAACCGGAGGAGGACCAAGAGTGTGATTGGGCATCGAGGGGACCCGAGAGG 3553
Qy 993 LysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3554 AAGGCGCTTCTCCAGTGTCTTTGTCCACATCCTGTCTCGGAC 3595

RESULT 8
US-09-023-905A-5
; Sequence 5, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023.905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-5

Alignment Scores:
Pred. No.: 9,06e-188 Length: 3456
Score: 2954.50 Matches: 622
Percent Similarity: 65.04% Conservative: 126
Best Local Similarity: 54.09% Mismatches: 239
Query Match: 56.03% Indels: 163
DB: Gaps: 24

US-09-914-042-1 (1-1006) x US-09-023-905A-5 (1-3456)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 61 ATGCCGATCATCATCTCCGTCTCCAGTTCCTCTCGGAGACGACGAGGATTACAAATTC 120
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaIle 40
Db 121 CCCAGACCTCGAGCTTCACCCCGCTGCTGAGAGTCCCGGACACACCGTCAATGTTCTG 180
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysValLysAla 60
```

Db 181 GAAAGGCTTTGGATCAGACCGAACTCTTTACAGAAAGGTCAAGAAATCTGTCAAAGCA 240  
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInThrGlnAlaLeuGlu 80  
Db 241 ATCTACAACTCGGGTCAAGAACATGTGCAGATGAAGAGATTAATGGACAGGCACTGGAC 300  
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100  
Db 301 AAGTTTGGCAACAACCTTCAAGCCGAGATAACTCTGATCTGGGAACAAGCCTTCTATCAAG 360  
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAsn 120  
Db 361 TTTTCTGGACCTTATCAAAGAGCTGCTCTCTCTCAAGAAACCTGCTCCAGAGCCTCAGC 420  
Qy 121 AsnIleLeuSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
Db 421 CACAACGTCATCTTCAACCTGGACTCTCTCTCAAGAGGAGATCTAAAGGGAGTGAAGGGG 480  
Qy 141 AspLeuLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
Db 481 GACCTTAAAGAGCCTTTCGCAAGGCTTGAAGAGACTATGAACCAAGTTCAACAAGATC 540  
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
Db 541 GAGAAGGAGAGAGAGAAACATGCCAAGCAGCACGGCATGATCCGCACAGAAATCACCGGC 600  
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
Db 601 GCAGAGATTGCAGAGAGATGGAGNAGGAGCGGAGGATCTTTCAGCTGCAGATGTGTAG 660  
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
Db 661 TACCTGATCAAGTCAATGAGATTAAAGACCAAGAGGAGTGGATCTCTCCAGAAATCTC 720  
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
Db 721 ATCAAGTATTATCATGCACGTGCAATTTCTCCAGGATGGCTTGAAACCTGCTGACAAAG 780  
Qy 241 LeuLysProSerIleGluThrSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
Db 781 TTGAAGCAGTATTATGAAATTAAGCAGCTGATCTTTATATATATAAACAAGACTCAGAT 840  
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
Db 841 GAGGAGAAAAACAGCTCACAGCTCTCAGAGACCTCATCAATCTTCTTACAGCTGGAC 900  
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
Db 901 CAGAAGGAGGATTCTCAGAGTAAGCAGAGC---GGGTACAGCATGTCACCAGCTGCAGGCG 957  
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
Db 958 AATAAGGAGTTGGCAGTGAGAGAGGGCTATCTCTTCAAGAGAGATGGAGATCCGT 1017  
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
Db 1018 AAGGTGTGCAGAGGAGGAGTGTCTCAGTGAATAATGTCATCTTCAACCATCTCTCATGCC 1077  
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360  
Db 1078 ACATCCAACAGCAGCCGGTGAGACTGAATCTGCTGACCTGCAGCTGCGAGTTAAACCCAGTGG 1137  
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 1138 GAGGATAAGAGTGTCTTGACCTCATCTCTCATATCAAGACATATCATTTTCCAGGACAG 1197  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
Db 1198 GACGAAACAGGAGTTGTGATATGATCTCGGTGCTGACTAATAGTAAGGAGGAGGCTCTG 1257  
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420

Db 1258 AACATGGCATTTCTGCTGGGAGCAGAGTCTGGAGATGACAGTTTG---GAGACTTGACC 1314  
Qy 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 1315 AAAGCCATCATCGAGGAGCTGCTGGCATTCCTGGAAACGAAGTCTGCTGTGACTGTGGG 1374  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 1375 GTTCAGAGCCCAATGGTTATCCACTAACCTCGGCATCTCTGACGTGATCGAGTGTCA 1434  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 1435 GGATTCACAGGGAATGGAGTCCATATTTCCGCATCCATCCATCGAGCTTGACAAA 1494  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1495 CTTGGAACCTCTGAACCTCTGCTGCTAAAGACCTGGGCAACAGTAGTTTCAACGAAATA 1554  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1555 TTAGAAGGGAATCTGCCGAGT---CCTTCAACAAGCCAGGCCCATCAAGTGACATGACC 1611  
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrTyrAlaArgLysLysHis 540  
Db 1612 GAGAGAGGAGTACATCAATCGAAGTACGTGGACACAGTTCGCTCGGGAACGCC 1671  
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
Db 1672 ACTACAGCCACAGCCAGACAGCGGCGACTTGTACAGGCGGTGAGAACCGAGACTTGATG 1731  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1732 GCTCTCATTCAGCTCTATGCAGATGGATGGAGCTAATGGATCTCTTCCCAAGACA--- 1788  
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
Db 1789 GGACAGGACCGGAGAGACAGCTCTGCACCTTGTCTGTCGACATCAGACACGAGACTTCC 1848  
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
Db 1849 CTGCACCTGTGGACTTCTTGTCCAAAACAGTGGGACTCTAGACAGACAGACGAGAGT 1908  
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
Db 1909 GGAACCGTGTCTCCATTTACTGTGCACATATGAGAGCCAGAGTGTCTCAAACTGTCTG 1968  
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
Db 1969 CTCAGGGGAAAACCGTCTATTGACCTGTTAATCAAAACGGGGAGACAGCATTTGATATC 2028  
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
Db 2029 GCCAGACGACTGAGAAATGTACAGTGTGAAGACTACTTGGTGGAGCAGCAGCGGAGG 2088  
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuLeuHisGluAspLeuAspGlu 700  
Db 2089 TTTAATCTCATGTGCATGTGGAGTATGAGTGAATCTGCGCTGAGGAGATTTGATGAG 2148  
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
Db 2149 AGTGACGATGACCTGGATGACAGCTTAGTCCAGTGAAGAGGAGCGTCTCTCTCGTCTC 2208  
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
Db 2209 CAGAGCTTC-----TGTCATTCGTCAGCGTGTCT---CCTCAG 2244  
Qy 741 AspAlaAlaAsnLeuAla-----LysGluLysGlnArg----- 751  
Db 2245 GAGAAGTTAAACCTCGCGGGGTATCTAGGACACAGGAGACAGAGACTGTCTCTATGGA 2304  
Qy 752 AlaPheMet----- 754  
Db 2305 GCCTTTGCCAACCCCGTCTACAGCACCTCCACCGAACCCTTCAGTGTCTCAGAGTGTGACAG 2364

```
Qy 755 ---ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPro----- 771
Db 2365 GGACCCACCATAGCCAGCAGACCCCTGCAGAAAGCTCCGTCCTGTGGCGCCGCCACCTCT 2424
Qy 771 ----- 771
Db 2425 CTGCGCTGGGATCTCAATCGAGTCGAGGAGGAGCTCCACTTCTCTAAGAGAGAGCT 2484
Qy 772 ProProAlaGlnProAlaProSerThrThrSerAlaProProLeuPro----- 788
Db 2485 CCTCCTCACCTCCCGGACACAAAGCGCACCCACTCAGATCCGCCCGCAGTCTGTCAG 2544
Qy 789 ---ProArgAsnValGlyValGlnThrAlaSerAlaAsnThrLeuTriplyThr 807
Db 2545 GGTCCGAGAGAAAGAGTGTAGTCCACCTCTCTGCAATTCGACATCCCGGCC 2604
Qy 808 AsnSerValSerValAspGlyGlySerArgGlnArgSerSerSerAsp----- 823
Db 2605 AAC-----AAGTTGAGGGAATCCAGCAGCAGCAAGCACTACGTCTATGACACAAAA 2658
Qy 824 -----ProProAlaValHisProProLeu----- 831
Db 2659 GCAACATTTGCCCCACGAGTTCTTCCCAAACTACCTCAAAAAGTGGCACTACGAAAGATT 2718
Qy 831 ----- 831
Db 2719 GACACAATCCACTCCCATCATAGTGACAAAGTCTGTGCTGTGATGTCTTGAGAACCCCA 2778
Qy 832 -----ProProLeuArgValThrSerThrAsnProLeuThr---ProThr 845
Db 2779 CAGGCCAGGATGCACCTCCACAGAGCCTCAGATACA-----ATAACAGACCCACT 2832
Qy 846 ProProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2833 GAACCTCCACCT-----AAAATTCACAGGTCCGACAGCAAGCATCCAGCCTGTGTGATGC 2886
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProPro-----LeuProPro 881
Db 2887 CCGAGAAACCGCACATCTCAGACTCTCTCCCAACCGCAACTATCAGATCTTCCCCCC 2946
Qy 882 GlnPro-----ProSerArgLeuProGlnLysLysProAlaProGlyThr 896
Db 2947 AAACCCCAATTGTCGATTTACCACCAAAA---CTCAGCTTCTGACCTGCCCCCGAG 3003
Qy 897 AspLysSerThrProLeuThrAsnLysGlyGln-----ProArgGlyProVal 912
Db 3004 CCTCAGCTTAAGGATCTTCCCCCTAAGCGCGAGATCAGTGATCTGCCATCCAAACCGGCC 3063
Qy 913 AspLeuSerAlaThrGluAla----- 919
Db 3064 GTGTGTTCTGCTGTGAGGCCACACAGAGCGAGTCAACGAGGAGGAACAGTCCGAG 3123
Qy 920 -----LeuGlyProLeu----- 923
Db 3124 CCCAGCTGACGAGACACAGTCATTACCCAGCAGGAGGAGCTCTCACCCGACAGGCC 3183
Qy 924 -----SerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
Db 3184 AGCAGGACACCAATGGAGCGCCCGCAGGAGGCTTTGGAAATGCCAGTCCCAATGCCAGCC 3243
Qy 938 Lys---SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysVal 956
Db 3244 AAAATTAAACAGATGACAAAGAACAAAGCGAGCGTGTGAACACCTCTATGATGGCCAG 3303
Qy 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGlu 976
Db 3304 CGAGACATGAGATGAGTGACTTTTGTGGAGGCGGAGTTATTAATTTGTACAGAGAG 3363
Qy 977 GluAspGlnGluTrrPrlleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Db 3364 GAAGACAGGAGTGTGGATCGGCACATAGAGGGTTCAGCCTGAAGAAAGAGGGGTCTTC 3423
```

```
Qy 997 ProValSerPheValHisPheIleAlaAsp 1006
Db 3424 CCATGTCCTTCGTGCACATTCGTCTGAC 3453
RESULT 9
US-09-023-905A-3
; Sequence 3, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4382
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (351)...(3803)
US-09-023-905A-3
Alignment Scores:
Pred. No.: 1,25e-187 Length: 4382
Score: 2954.50 Matches: 622
Percent Similarity: 65.04% Conservative: 126
Best Local Similarity: 54.09% Mismatches: 239
Query Match: 56.03% Indels: 163
DB: 4 Gaps: 24
US-09-914-042-1 (1-1006) x US-09-023-905A-3 (1-4382)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 411 ATGCCGATCAGATCTCGTGTCCGAGTTCTCTCGGAGACGACGAGGATTAACAATCC 470
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAla 40
Db 471 CCCAGACCTCGAGCTTCACCACCGCTGCGAGAGCTGCCGGAACACACGGTCAATGTTCTG 530
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60
Db 531 GAAGAGGCTTTGGATCAGGACCGAACTGCTTTACAGAAAGTCAAGAAATCTGTCAAAGCA 590
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 591 ATCTACAACTCGGTCAAGAACATGTGAGATGAGAGAAATTTATGGACGCACTGGAC 650
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLys 100
Db 651 AAGTTTGGCACTTCATCAGCCGAGATACTCTGATCTGGAACACGCTTCATCAAG 710
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 711 TTTTCTGACTTATCAAGAGCTGCTCTCTCAAGAACCTGCTCCAGAGCTCAGC 770
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 771 CACAGCTCATCTTCAACCTGGACTCTCTGCTCAAGAGAGATCTTAAGGGAGTGAAGGG 830
Qy 141 AspLeuLysLysProPheAspLysAlaTriplyAspTyrGluThrLysIleThrLysIle 160
```





```
Qy 824 -----ProProAlaValHisProProLeu----- 831
Db 3009 GCAACATTGGCCACGAGTTCTTCCCAAACTACCTCAAAAGTGGCACTAGCAAGATT 3068
Qy 831 ----- 831
Db 3069 GACAAATCCACCTCCCATCAGTGCAGCAAGTCTGGTCTCTGATGTGTTTCAGAAACCCCA 3128
Qy 832 -----ProProLeuArgValThrSerThrAsnProLeuThr-----ProThr 845
Db 3129 CAGGCCAGGATGACCTCTCCACGACGCTCAGATACA-----ATAACGAGACCCACT 3182
Qy 846 ProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 3183 GAACCTCACCT-----AAATTCACAGTTCGACAGAACGATCCACGCTGTGGATGTC 3236
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProPro-----LeuProPro 881
Db 3237 CCGCAGAAACCGCACATCTCAGACCTTCTCCCAACCGCAACTATCAGATCTTCCGCCCC 3296
Qy 882 GlnPro-----ProSerArgLeuProGlnLysLysProAlaProGlyThr 896
Db 3297 AAACCCCAATTTGCGATTATCCACCAAAA---CCTCAGCTTTCTGACCTGCCCCCGAAG 3353
Qy 897 AspLysSerThrProLeuThrAsnLysGlyGln-----ProArgGlyProVal 912
Db 3354 CCTCAGCTTAGGATCTTCCCTTAAGCGCGAGATCAGTGATCTGCCATCCAAACCGGCC 3413
Qy 913 AspLeuSerAlaThrGluAla----- 919
Db 3414 GTGTGTTCTGCTCTGAGGCCACACAGAGGCGAGTCAACGCGAGGAGGAACAGTCCGAAG 3473
Qy 920 -----LeuGlyProLeu----- 923
Db 3474 CCCAGCTGACGGAGACACAGTCATTCAGCCAGCAGGAGGACTCTCACCCGACAGGCC 3533
Qy 924 -----SerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
Db 3534 AGCAGGACACCAATGGAGCGCCCGCAGGAGCTTGGAAATGCCAGTCCCAATGCCAGCC 3593
Qy 938 Lys---SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysVal 956
Db 3594 AAATAATACACAGTAGCAAGCAAGAAACGAGCGTGTGAAACCATCTATGATGTCAG 3653
Qy 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGlu 976
Db 3654 GCAGACATGACGATGAGCTGACTTTTGTGGAGGCGGAGGTATATAATTGTTCACAGGAG 3713
Qy 977 GluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Db 3714 GAAGACCAAGGAGTGGTGGATCGGGCACATAGGGGTGACGCTTGAAAGGAAGGGGTCTTC 3773
Qy 997 ProValSerPheValHisPheIleAlaAsp 1006
Db 3774 CCAATGCTCTTCGTGCACATTCGTGCAGAC 3803
```

## RESULT 10

US-09-620-312D-155  
; Sequence 155, Application US/09620312D  
; Patent No. 6569662

## GENERAL INFORMATION:

APPLICANT: Tang, X. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunqing  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John fillinghast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt\_FL\_genes version 1.0  
SEQ ID NO 155  
LENGTH: 6990  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (61)..(2397)

## US-09-620-312D-155

## Alignment Scores:

Pred. No.: 3,28e-28 Length: 6990  
Score: 571.50 Matches: 196  
Percent Similarity: 41.68% Conservative: 137  
Best Local Similarity: 24.53% Mismatches: 305  
Query Match: 10.84% Indels: 161  
DB: 4 Gaps: 25

## US-09-914-042-1 (1-1006) x US-09-620-312D-155 (1-6990)

```
Qy 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyLysMetLysLys 56
Db 115 GCAGCTTTGGAAGAAGTAGAAGGTGATGGCAGCAATTGGAACTA---AAACTTGATAAG 171
Qy 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyThr 76
Db 172 CTTGTGAACCTT-----GGGATTGCAATGATTGATCTACTGGAAAGACCTTTTGT 219
Qy 77 GlnAlaLeuGluLysPheGlyAsnCysValCysArgAsp----- 90
Db 220 GTTCAATAAACAGTTCATGAATGGG-----ATTCGAGACCTGGCCCATTTCTAGT 273
Qy 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
Db 274 AATGATGCTGCTGTGAGACAAAGTTTGACCAAGTTTCTGACAGTCTTCAAGAAATGATA 333
Qy 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnIleIleSerPheProLeuAspSer 129
Db 334 AATTTTCACACAATCCTGTTGCCAAACTCAGAGATCAATTTTAAGGCACAGCTTCAGTAAAC 393
Qy 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPheAspLysAla 149
Db 394 TTTGTTAAAGAAAGATCTTAGAAATTCAAA---GATGCCAAGAACGCAATTTGAAAGATC 450
Qy 150 TrpLysAspTyThrLysIleThrLys-----IleGluLysGluLysGlu 166
Db 451 AGTGAAGAAAGAAATCGCTTAGTAAAAAATGCCCAAGTACAAAGAAACAAACACAT 510
Qy 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGlu 186
Db 511 GAAGTTGAA-----GAAGCCCAACCAACATT 534
Qy 187 MetGluLysGluArgArgPhePheGlnLeuGlnMetCysGluTyThrLeuLysValAsn 206
Db 535 CTGACAGCAACAAAGAAATGTTTCCGACACATAGCCCTCGAATATGCTTCCAGATTAAAT 594
Qy 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyPheHisAla 226
```



Db 595 GTTCTTCAATCAAAAGGAGATCAGAAATCTTAAATCAATGTTGTCTATTTATGTATGCC 654  
Qy 227 GlnCysAsnPhenPheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGlu 246  
Db 655 CATTTGGCCCTTTTCATCAAGAGATATGATCTGTTAGTGAATCTTGGACCTTACATGAAG 714  
Qy 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeu 266  
Db 715 GATCTTGTGTGACAGTGTGATCTGTTGATGTCAGCAAGAGGAGAAAGAGAAATG 774  
Qy 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGln 286  
Db 775 GAGCAAAACATTC-----ACCATTCAACAAAGGATTTCTCCAGT 816  
Qy 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306  
Db 817 GATGATTTCTAAGTTAGATATAACGTA-----GATGCTGCAAAATGGCATA 861  
Qy 307 GluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArg 326  
Db 862 GTTATGGAAGGATATCTGTTCAACAGCGCCAGCAATGCCCTTCAAAACTTGGAAACAGCGC 921  
Qy 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346  
Db 922 TCGTCTTCAATACAGAAATATCAGTTGTTTACCAGAAAAATTTAAGGATTAATCCGACT 981  
Qy 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGluLys 363  
Db 982 GTGGTAGTTGAAGACCTCAGGCTTTGCACAGTGAACATTTGTGAAGACATAGAGCGACA 1041  
Qy 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383  
Db 1042 TTCTGCTTTGAGGTGCTCGCAACAAAAAGTTGCATGCTCCAGCGAGATTTCCGAAAAG 1101  
Qy 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403  
Db 1102 CTGCCAGGCGATGATTAAGCTGTTCCAGCACGATTTGTCTACTGTCTATAGAGAG--- 1158  
Qy 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412  
Db 1159 ---AAGGTGATGATCAGAGAAGCTGGATAAGAAATCATCTCCATCCACAGGAAGCCTA 1215  
Qy 413 -----AsnAlaValGlnGluLeuThrLys-----GluIleIleSerGluVal 427  
Db 1216 GATTCGAAATAGCTCCAAAGAGAAATTAATTGAAGGAGAAAGTGCCTTCAGCGGCTC 1275  
Qy 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447  
Db 1276 CAGTGATCCCTGGCAATGCCAGCTGTTGTGACTGTGGCTGGCAGATCCACGGTGGGCC 1335  
Qy 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467  
Db 1336 AGCATCAACCTGGGCATCACCTTGTGTATCGAGTCTCCGGAATTCACGGAGCCTTGGG 1395  
Qy 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487  
Db 1396 GTTCATTTTCAAAAGTACGATCTTTAACTTTAGACACTGGGAGCCAGAACTTTTAAAG 1455  
Qy 488 LeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAla 507  
Db 1456 CTTATGTGTGAGTGGGAGATGATGTTTATAATCGAGTTTATGAAGCTTAATGTGGAATA 1515  
Qy 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527  
Db 1516 ATGGCAATAAGAAACCCCAACAGGA---CAAACAGAGGAGAGGAGGCATATATCAGA 1572  
Qy 528 AlaLysTyrIleGluArgTyrAla----- 536  
Db 1573 GCAAAATATGTGAGAGGAAATTTTGGGATAAATATTCTATATCATTTATCCTCCTCGAG 1632  
Qy 537 ---ArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCys----- 551  
Db 1633 CAGCAAAAAAAGTTTGTCTCTTAAAGTTCTTGAAGAAAAAGAGGCTGAGCATTTCTAAATTT 1692

Qy 552 -----GluAlaValLysThrArgAspIlePhe 560  
Db 1693 GGGCAGCGGGACCAAGTCAGAGCATCTGCCCAAGTTCAGTCAGAAGTAATGAC---AGT 1749  
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
Db 1750 GGAATTCAGCAGAGCTCTGATGATGGAAGAGAAATCTTTACCTCCACGGTGTGAGCCCAAT 1809  
Qy 581 Gly---HisGluProAsp----- 585  
Db 1810 AGTTTATATGAGCTGAAGGAGAAAGGCAAGATTTCTTATGTTTCTTGTGACTCGAAACAT 1869  
Qy 585 ----- 585  
Db 1870 CTTAATCCAGGACTTCAGCTTTATAGGGCGTCATATGAAAAAACCTTCTCTAAATGGCT 1929  
Qy 586 GluThrAlaLeuHisLeuAla----- 592  
Db 1930 GAGGCTTTGGCTCATGCTGGTCAGACGTCGAACCTGGGCAATTCAGAGGAAAAACAAGCGACA 1989  
Qy 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610  
Db 1990 CCACCTTATTCAGGCTGATTTAGGGGCTCTTTGGTGACGCTGTGAGTTCTCTCTACAGAT 2049  
Qy 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630  
Db 2050 GGTGCTAATGTCAACCAAGAGATGTCCAAGGGCGGGGACCATTCACCATCCACCGCTC 2109  
Qy 631 ThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAla 650  
Db 2110 TTAGGGCACACAGGCGAGTATGTTTATTCCTAAACAGAGGTGCCAATCAACATGCCACT 2169  
Qy 651 AsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670  
Db 2170 GATGAGAGGAGAAAGACCTTTTGACATAGCTGTGGAAGGAGCCCAATTCATATAGTTC 2229  
Qy 671 GluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690  
Db 2230 ACCTTGTTCCTTAGCA----- 2247  
Qy 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710  
Db 2248 -----AGAATGAATGAAGAGATGCGGAATCAGAGGAGACTTTATGGA-----CAG 2292  
Qy 711 ProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729  
Db 2293 CAGGTGATGAACCTTATCAGACATATTTCTGTGATTTTTCCTCAATGGCATCCCAAT 2349

## RESULT 11

US-09-270-767-1629/c  
; Sequence 1629, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1629  
; LENGTH: 1090  
; TYPE: DNA  
; ORGANISM: *Drosophila melanogaster*  
US-09-270-767-1629  
Alignment Scores:  
Pred. No.: 2,12e-21 Length: 1090  
Score: 453.00 Matches: 84  
Percent Similarity: 77.55% Conservative: 30  
Best Local Similarity: 57.14% Mismatches: 33  
Query Match: 8.59% Indels: 0

```
DB: 4 Gaps: 0
US-09-914-042-1 (1-1006) x US-09-270-767-1629 (1-1090)
Qy 2 ProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAlaPro 21
Db 443 CCATCCCTGATTGCGGTGCGAGTTTGTGGAGGAGACGCGCTCGGATTACAGCTCGCCC 384
Qy 22 ThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIleGlu 41
Db 383 ACCACCTCCACCTTCGCGCTCGAGGATGCCGAGTCCCGCACCACGATTTGGCGTCTTGGAA 324
Qy 42 GluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAlaIle 61
Db 323 GAGATTGGAGTTCCAGCGGAGGGTCTAACTAAGCTGAAAAAGCGGTCAAGCTATC 264
Qy 62 AsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGluIys 81
Db 263 CACAACCTCAGGAACACCCATGTGACATGAGATGTTTATGGTCGAGCTCTGAGGCGA 204
Qy 82 PheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuIysPhe 101
Db 203 CTGGCGGCAAGGTTATTGAGCAGGATGAGCGCGACATCGCGCGCGTTCCTTAAGTTC 144
Qy 102 SerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsn 121
Db 143 AGCGTTGTCCACCAAGAGCTTAGCGCGCTGATGAAGACCTGTATGCAAAAACATCAACAAC 84
Qy 122 IleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGlyAsp 141
Db 83 ATTGTGATGTTCCGGTGGACTCAATGCTGAAGAGCGAGCTCGCGGGCGTGAAGGGGAC 24
Qy 142 LeuLysLysProPheAspLys 148
Db 23 ATGAAGCGCCATTCGACAAG 3
RESULT 12
US-09-270-767-16911/c
; Sequence 16911, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16911
; LENGTH: 1090
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16911
Alignment Scores:
Pred. No.: 2,128-21 Length: 1090
Score: 453.00 Matches: 84
Percent Similarity: 77.55% Conservative: 30
Best Local Similarity: 57.14% Mismatches: 33
Query Match: 8.59% Indels: 0
DB: 4 Gaps: 0
US-09-914-042-1 (1-1006) x US-09-270-767-16911 (1-1090)
Qy 2 ProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAlaPro 21
Db 443 CCATCCCTGATTGCGGTGCGAGTTTGTGGAGGAGACGCGCTCGGATTACAGCTCGCCC 384
Qy 22 ThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIleGlu 41
Db 383 ACCACCTCCACCTTCGCGCTCGAGGATGCCGAGTCCCGCACCACGATTTGGCGTCTTGGAA 324
Qy 42 GluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAlaIle 61
```

```
Db 323 GAGATTGGAGTTCCGACCGGAGGGTCTAACTAAGCTGAAAAAGCGGTCAAGCTATC 264
Qy 62 AsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGluIys 81
Db 263 CACAACCTCAGGAACACCCATGTGACATGAGATGTTTATGGTCGAGCTCTGAGGCGA 204
Qy 82 PheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuIysPhe 101
Db 203 CTGGCGGCAAGGTTATTGAGCAGGATGAGCGCGACATCGCGCGCGCTTCCTTAAGTTC 144
Qy 102 SerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsn 121
Db 143 AGCGTTGTCCACCAAGAGCTTAGCGCGCTGATGAAGACCTGTATGCAAAAACATCAACAAC 84
Qy 122 IleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGlyAsp 141
Db 83 ATTGTGATGTTCCGGTGGACTCAATGCTGAAGAGCGAGCTCGCGGGCGTGAAGGGGAC 24
Qy 142 LeuLysLysProPheAspLys 148
Db 23 ATGAAGCGCCATTCGACAAG 3
RESULT 13
US-09-828-303-11
; Sequence 11, Application US/09828303
; Patent No. 6677504
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND
; FILE REFERENCE: 16313-0030
; CURRENT APPLICATION NUMBER: US/09/828,303
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2041
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-303-11
Alignment Scores:
Pred. No.: 2e-19 Length: 2041
Score: 429.00 Matches: 154
Percent Similarity: 39.30% Conservative: 103
Best Local Similarity: 23.55% Mismatches: 217
Query Match: 8.14% Indels: 180
DB: 4 Gaps: 23
US-09-914-042-1 (1-1006) x US-09-828-303-11 (1-2041)
Qy 134 AspLeuLysGlyValIysGlyAspLeuLysLysProPheAspLysAlaTrpLysAspTyr 153
Db 14 GACTTCCGAATGCAAAATCACA-----TTTGAGCAGGCTCGATTCAATTG 61
Qy 154 GluThrLysIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHisGlyMet 173
Db 62 ATGACGACCTACCAATAGTAGGAGCAAAAAGAGTTTCGAG----- 103
Qy 174 IleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLysGluArgPhe 193
Db 104 TTCCTTGAGCGGTGAGTGGTACA-----ATGGATGCACATCTCAGGTAC 148
Qy 194 PheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIleLysGly 213
Db 149 TTC-----AAGCAGGCG 160
```



Db 695 AGCAGCATGGAAGAGCGTTACGATTTTATCCGCGCCAAAGTATGTTGCCAAGCGCTATGTC 636  
Qy ArglylsYsHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThr 556  
Db ATGCGCAGCTGCTCGGATGACAAAGCAGCTTACGGTGCAGCTTAGAGCAGCGCTGTGCTCAAC 576  
Qy ArgAspIlePheGlyLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIle 576  
Db CCGGACATGAGTCACTGCTCAGGTGTGGCGGAGGAGCGGATCTCAGCTGCTGCTG 516  
Qy ProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVal 596  
Db CCC-----AGCTCTGATCGCGAGAGCGGCCCTCCATCTGCGCGTGTGCGCGAG 465  
Qy AspArgThrSerLeuHisIleValAspPheLeuValGlnAsn----- 610  
Db ATGGGATCCAGCTGCACATAGTACATTTCTCATCCAAATATATGCCGCCCTAAAGGCTT 405  
Qy -----SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAla 624  
Db AACAAAGTACCAATCCGCGCGCTCTGTGGAT---GTGACGGGAAAG---AATACGGCG 351  
Qy LeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLys 644  
Db TTGCATTGTGGCTCTGCAGATCGGAGGAGTGCATGAAGTCTGCTCGCTCTGGA 291  
Qy AlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeu 664  
Db CCGGACTATGACTCAAGAAGCTCAGAACAAACAGCTCTGGATATTTGCCAAAGAAATG 231  
Qy LysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHis 684  
Db GGACACATAGCTGACGGAGGCTTATGAATGTGCATTAAGCGGGAGAGCGCCTTC 171  
Qy ValHisValGluTyrGluTrpArgLeuLeuHisGluAsp-----LeuAspGluSerAsp 702  
Db GACCATCAACACTGACTGGAATTCGCAAAAGAGGACGGTTCCACCGATTTCAGCGAC 111  
Qy AspAsp-----MetAspGluLysLeuGlnProSer 712  
Db GATGAACAGCATCATCGATGAGCGCTCCAGTTCTAGT 75

## RESULT 15

US-09-914-042-35  
; Sequence 35, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John, Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 35  
; LENGTH: 2119  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (520)..(1614)  
US-09-620-312D-35  
Alignment Scores:  
Pred. No.: 4e-12 Length: 2119  
Score: 320.00 Matches: 108  
Percent Similarity: 38.07% Conservative: 50  
Best Local Similarity: 26.02% Mismatches: 159  
Query Match: 6.07% Indels: 98  
DB: 4 Gaps: 10

US-09-914-042-1 (1-1006) x US-09-620-312D-35 (1-2119)  
Qy 310 GlySerLeuTyrLysLysSer---AspGlyIleArgLysValTrpGlnLysArgLysCys 328  
Db 409 GGGATCTCTGCTAAAGCGGAGCGCAAGTCCCTGTAACAAGAGAGTGAAGAAGATGTG 468  
Qy 329 SerVal---LysAsnGlyPheLeuThrIleSer----- 338  
Db 469 ACGTCTGTGCACACGGGCTCTCACCTATCACCCAGCTGCATGATTATCATGCAGAAC 528  
Qy 339 ---HisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
Db 529 ATCCACGGC-----AAGGAGATTGACCTGCTGCGGCAACACGCTGAAA 570  
Qy 357 ----- 357  
Db 571 GTGCGAGGAAGCGCTGCCCGAGCCACACCTGCCACAGCCCCCGGCCAGCCCCCGT 630  
Qy 358 -----ThrAsnProGlu 361  
Db 631 GCCAACGGGCTGCTCGTGGAGCGGAGTAACACACAGCTGGTGGTGGGGCAGAGCGCAG 690  
Qy 362 GluLysLysCysPheAspLeuIleSer---HisAspArgThrTyrHisPheGlnAlaGlu 380  
Db 691 GAGTCGTTTGAATTTGTGGTGTCTCCTCACTGGCAGACGCTGGCACTTCGAGGCTTCA 750  
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu 400  
Db 751 ACGCGGAGGAGCGGAGCTGTGGGTTTCAGAGTGTGCAGGCCCCAGATCCTTTGCCAGCCTG 810  
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
Db 811 CAAGCTGCCGACGTGCCAAGGACAGACTCGACTGGGGAACACAGACGCGAGCTCTGGCT 870  
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
Db 871 -----GTGACGGCGCTCCGACCGCTCCGCGCAACAGACTTTTGTATCGACTCGCAT 921  
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
Db 922 GCACCAATCCAGACTGGGCCAGCCTGAACCTGGGTCGCCCTGTGATGATTTGAGTGTCA 981  
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
Db 982 GGCATCCACCGACACTGGGGGCTCACCTGTCCCGGGTGGCGCTCCCTTGACTCGATGAC 1041  
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
Db 1042 TGGCGCGCTGAGCTGCTGGCTGTCATGACTGCCATGGCAATGCGCTCGCCCAACAGCGTC 1101  
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
Db 1102 TGGGAGGGGGCTTTGGGTGGCTACTCC-----AAGCCAGGCGCTGTGCTCGAGAG 1155



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 4, 2005, 08:27:30 ; Search time 1229 Seconds  
(without alignments)  
4845.614 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVFSEVAETHDYKA.....DGDPRKGAPFVSFVFIAD 1006

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool\_b/US09914042/runat\_04082005\_090631\_2104/app.query.fasta\_1.1159  
-DB=N\_Geneseq -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORES=spect -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914042 @CGN 1 1 615 @runat\_04082005\_090631\_2104 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseq16Dec04:\*
- 2: Geneseq1980s:\*
- 3: Geneseq1990s:\*
- 4: Geneseq2000s:\*
- 5: Geneseq2001as:\*
- 6: Geneseq2001bs:\*
- 7: Geneseq2002as:\*
- 8: Geneseq2002bs:\*
- 9: Geneseq2003as:\*
- 10: Geneseq2003bs:\*
- 11: Geneseq2003ds:\*
- 12: Geneseq2004as:\*
- 13: Geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3874	73.5	5954	2	AAV59105
2	3417	64.8	4328	2	AAQ28758 Zebrafish
3	3155	59.8	4595	2	AAV59106 Zebrafish
5	3029	57.4	3775	10	Add71214 Human int
	3024.5	57.4	3946	10	Add71207 Human int

6	3013.5	57.1	5330	2	AAV59103
7	3006	57.0	3974	12	ADJ51008
8	2954.5	56.0	4382	2	AAV59104
9	2852	54.1	5947	13	ADS34277
10	2556.5	48.5	4997	13	ADS10088
11	2555.5	48.5	5033	4	AAI59224
12	2455.5	46.8	4870	13	ADS34272
13	2410	45.7	5475	13	ADS34273
14	2099.5	39.6	3812	10	ADBG2827
15	2089.5	39.6	4050	10	ADK81946
16	1803.5	34.2	3830	4	ABL04309
17	1801.5	34.2	3436	4	ABK43447
18	1801.5	34.2	3436	12	ADI53834
19	1759.5	33.4	3413	4	ABK43768
20	1759.5	33.4	3413	12	ADI54155
21	1583	30.0	6688	4	ABL04308
22	1307.5	24.8	3123	10	ADBG2160
23	802	15.2	1435	4	AAS33079
24	766.5	14.5	544	13	ADQ79413
25	695	13.2	563	4	AAS58478
26	611	11.6	2563	6	ABN59873
27	587	11.1	1224	8	ACA57215
28	581.5	11.0	5319	6	ABN83974
29	573.5	10.9	3926	10	ADD71213
30	571.5	10.8	6990	4	AAI58278
31	571.5	10.8	6990	5	ADQ98485
32	571.5	10.8	6990	9	ADB48245
33	570.5	10.8	2484	10	ABX70672
34	561	10.6	2737	6	ABK88008
35	558	10.6	5421	12	ADQ25213
36	552.5	10.5	6979	5	AAS81429
37	534	10.1	2877	13	ADS09828
38	532.5	10.1	2505	6	ABK71808
39	532	10.1	747	6	ABK71808
40	509.5	9.7	2494	6	ABK83644
41	509.5	9.7	2494	13	ADS34261
42	509.5	9.7	2498	13	ADS34262
43	509	9.7	2223	13	ADS34264
44	509	9.7	6480	6	ABK83626
45	493.5	9.4	7027	4	AAI60064

ALIGNMENTS

RESULT 1

AAV59105  
ID AAV59105 standard; cDNA; 5954 BP.

XX AC AAV59105;

DT 17-OCT-2003 (revised)  
DT 20-NOV-1998 (first entry)

DE Zebrafish differentiation enhancing factor 2 gene.

XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
XX C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
XX KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
XX KW nervous system.

XX OS Danio rerio.

XX FH Key Location/Qualifiers  
XX CD5 433..3381  
XX FT /\*tag= a  
XX FT /product= "DEF2 protein"

XX WO9836065-A1.

XX PD 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US002724.

XX

PR 14-FEB-1997; 97US-0038191P.  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;  
 XX WPI; 1998-467173/40.  
 XX P-PSDB; AAW77288.  
 XX New nucleic acid encoding differentiation enhancing factor - used  
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
 PT tumours and neurological disease.  
 XX Claim 7; Fig 14; 203pp; English.  
 XX The differentiation enhancing factors (DEF), comprise at least one each  
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
 CC are mediators of SH3-domain dependent signalling and may be involved in  
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
 CC endocytosis or adhesion, migration, proliferation and differentiation of  
 CC cells. Typical applications of DEF and agents that modulate interaction of  
 CC between the protein and it's ligand, or of nucleic acid expressing them,  
 CC are treatment of hyperplastic and neoplastic disease (a wide range of  
 CC solid tumours and leukaemias) including metastases; for in vitro  
 CC induction of differentiation of neural crest cells to neurons, glial  
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
 CC the nervous system (e.g. treatment of traumatic injury, stroke,  
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 5954 BP; 1731 A; 1382 C; 1372 G; 1469 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4,89e-162 Length: 5954  
 Score: 3874.00 Matches: 755  
 Percent Similarity: 83.12% Conservative: 92  
 Best Local Similarity: 74.09% Mismatches: 122  
 Query Match: 73.47% Indels: 50  
 DB: 2 Gaps: 13  
 US-09-914-042-1 (1-1006) x AAV59105 (1-5954)  
 QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
 DB 433 ATGCTGACCCAGATTAACATGCGGAGTTTGTACCGAGACAAATGAAGATTATAAATCG 492  
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaTle 40  
 DB 493 CCCACCGCTCAAACTTCAACACCAAGATGACTCACTGCAGGAACACAGTATCCGCACTG 552  
 QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
 DB 553 GAGGAGGCCCTGGATGTGACCGCAGTGTCTTTTACAAGATGAAGAGTCAAGTAAAGGCT 612  
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyrThrGlnAlaLeuGlu 80  
 DB 613 ATTTACGGCTCGGGTCTGGCTCATGTGAGAAATGAGGAGCATACACTCAAGCTCTGGAG 672  
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
 DB 673 AAGTTGGAGAGAACTGTGTGTACAGAGATGACCCGGACCTTGGGATCAGCTTCTCTGAAG 732  
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 DB 733 TTCTCCGCTTCCACCAAGAGCTCAGGCACCTCTTCAAGAACTCTTTCAAGAACATGAAT 792  
 QY 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140  
 DB 793 AATATCAVTAACCTTCCCAATTGACAGTCTGTGTAAGGGAGATCTGAAAGGGGTTAAAGGG 852  
 QY 141 AspLeuLysLysProPheAspLysAlaIleLysAspTyrGluThrLysIleThrLysIle 160

DB 853 GATCTCAAGAAAGCCCTTCGATTAAGCCTGGAAAGACTACGAGACTAAAGTCTCTAAATA 912  
 QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
 DB 913 GAGAAGGAGAAATAAGAGACGCGCCGAGCAGCGAATGATCCGAGCGAGATCAGCGGA 972  
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200  
 DB 973 GCAGAGATAGCAGAAGAGATGGAAAAAGAGCGCGCTTCTTCCAGCTTCAGATGTGTAG 1032  
 QY 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
 DB 1033 TACCTCTCTCAAGTCAATGAATCAAGATCAAAAAAGGTGTCCAGCTGTCTCCAGATCTC 1092  
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValIleSer 240  
 DB 1093 ATCAAAATACCTTCCACGCACACAGTGCACCTCTTTTCCAGATGGTCTCAAAAGCGGTGCACAAC 1152  
 QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 DB 1153 CTCAAAACCTCAATAGAAAAACTGCGCACAGACTTGCACTCGATCAAAACACAGGTACAGAT 1212  
 QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
 DB 1213 GAAGAACGCACACAGCTTACCAGTTACGGATGTGTAAAAACTGTCTCTCAAGTGGAG 1272  
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 DB 1273 CAGAAGGAGGACTCTCAGGTAGACAGAGCGCCACCTACAGTCTGCACACCGCCAGGGC 1332  
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
 DB 1333 AACAAAGAGCATGGGACTGAGCGCAGCGGCAACCTTTTACAAAGAGAGTACGGGCTGCGG 1392  
 QY 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
 DB 1393 AAGTGTGGCAGAGAGAAAGTGCACAGTAAAGAAATGGATATTTGACCACTCTCACATGGG 1452  
 QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
 DB 1453 ACGGCAAAACAGACCTCCCGCCAAACTCAATCTTCTCACTGTCTCAGGTGAAGACAAACCCA 1512  
 QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
 DB 1513 GAGGAGAAGAAAAAGTTTGTACCTCATCTCACATGACAGAAATATCATTTTCCAGGCAGAA 1572  
 QY 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
 DB 1573 GATGAGCCAGAGTGTCAAAATATGGATCTCAGTGTGCGAGAACAGTAAAGAAGAGCGGCTC 1632  
 QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420  
 DB 1633 AACACGCTTCAAGGGCGCACAGCATGTTGGTGAATAAATCAATTTGTGAGAGGTCCACC 1692  
 QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
 DB 1693 AAGGCCATCTCTGGGAGAGGTGAAGCGGATGCGGGGAACGATGTCTGCTGCGACTCGCGT 1752  
 QY 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460  
 DB 1753 GCTCCCGGCCCAATGGCTCTCCACCAACTGCGGCATCTCGACTGCATGCATGAGTGTTCG 1812  
 QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
 DB 1813 GGGATCCACAGAGAGCTGGGCGTCCATTACTTCCCGAATCCAGTCCCTCACACTCGACGTC 1872  
 QY 481 LeuGlyThrSerGluLeuLeuAlaIleAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
 DB 1873 CTCAGACCTCCGAGCTCTTGTCTGGCCAAAGAACGTGGGAAATGCTGGCTTCAATGAGATC 1932  
 QY 501 MetGluCysLysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520



```
Db 1933 ATGGAGGCTGTCTGACGCGAAGATGTGATCAACCCGATCCAGCCAGTGACATGCAG 1992
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysHis 540
Db 1993 GCGAGGAAGGACTTTATCATGCGCAAAATACACAGAGAAACGCTTCGCTCGTAAGAAGTGT 2052
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 2053 CCAGACGCACTGTGGAAGCTGCACACGCTGTGTGATGCTGTAAGCCCGGACATTTTC 2112
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2113 TCTCTCATCAGGCTATGCTGAAGAGTGGATCTGATGGACCCATCTCTCTGCTAAT 2172
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2173 GCACATGAACAAGGTGAGACGGCTCTTCATCTGGCGGTGAGACTGGTGGACAGAACTTCC 2232
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuLeuAspLysGlnThrGlyLys 620
Db 2233 CTACACATCATCGACTTCTCCACCAAAACAGTTTAAACCTGGATTAACCAACCGCTAAA 2292
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2293 GGAAGCAGAGCTCTGCATCTACTGCTGCTGACGACAAACGCGAGTGCTCAAACTGCTG 2352
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2353 CTCAGAGAAAGCCCTCCATGATATATGCTTAATGAAGCTGGAGACACCCCGTTGGACATC 2412
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2413 GCCAGCGCACTCAACATCTGCAGTGTGAGGAACCTGCTGAACACGAGCTCTTTCAGGGAAG 2472
Qy 681 PheAsnSerHisValHisValGluTyrGluThrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2473 TTCAATGCTCATGTCATGTGAGTATGAGTGGAGACTTCAGCTGAAGAGCTTGAGCCAG 2532
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2533 AGTGATGAAGATCTGATGAGAGTCGAGTCTCT-----CACCGCGGGATGAGCGGCC 2586
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2587 ATCAGCTGCTACACACCGCGGAGTAATCTCCCTTCAGCTGAGTCCAGCCAGCGCTGAGCGGA 2646
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2647 GACGCTCAGAGACCTGGTTAAAGACAAACAGCGC---TTGTGCCAAACCTGGTCAACAAAT 2703
Qy 761 GluThrTyrGlyAlaLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2704 GAAACCTACGGGACCATTAATACACAGCTCACCC-----GTACGCGCTG 2748
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2749 TCCTCTTCTGCTCCACCTCTACACCCCGAAACCTA-----GTTACGCGCTGCTCTT 2802
Qy 801 AlaAsnThrLeu-----TrrLysThrAsnSerValSerValAspGly 815
Db 2803 CGAGACTGACTCAAGGATCTCCCGGCTGGAGCCTGCTCTCTGGATCTGAGCGGC--- 2859
Qy 816 SerArgGlnArgSerSerAspProAlaValHisProProLeuProProLeuArg 835
Db 2860 ---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTGCGCCCTCCCTTACGG 2916
Qy 836 ValThrSerThrAsnProLeuThrProThr---ProProProProAlaLysThrPro 854
Db 2917 GTCACTTCACCTCTCTAATGCCCAGCGGTGCTGCTCTCTCTCTGCTAAAGCTACT 2976
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProGlyIleSer 872
Db 2977 GGTATGATGGAGACCATGAATATGCAACCCCAACCCGAGCGGGGCTCTCTGGACAGAAC 3036
```

```
Qy 873 GlnIleArgProProProLeuProGlnProProSerArgLeuProGlnLysLysPro 892
Db 3037 ATCAACCGG----- 3045
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 908
Db 3046 GCTCAAGTGGCGGACAAAACCTTCAGCAAAAGCACACTGATGCGCTCC----- 3093
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927
Db 3094 ---GGATCCATCGAGAGACCACTTAAGAAGTCCCGAGGAGGCCACAAAACACCACTGGT 3150
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947
Db 3151 CAAATCTGCTGCGGACCCACATGCCAGGAAA-----ACGTATTTGAAGCCGAAG 3201
Qy 948 ArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGlu 967
Db 3202 CGTGTGAAGGCCATGTATACTGTGTGCCGATATATCCAGACGAGCTGACCTTCTCTGAG 3261
Qy 968 GlyAspValIleIleValAspGlyGluGluAspGlnGluTrrPrrIleGlyHisIleAsp 987
Db 3262 GGAGAGCTTATCGTGTGGTGGATGGAGAGGAGGACGAGAGTGGTGGCTGGGCCACATTGAG 3321
Qy 988 GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3322 GGAGAGCCAATGAGAAGAGGAGCGTTTCTGTCTGACGTTTGTACAGTTTCATTATGGAC 3378

RESULT 2
AAQ28758
ID AAQ28758 standard; DNA; 4328 BP.
AC AAQ28758;
XX
XX 25-MAR-2003 (revised)
DT 25-FEB-1993 (first entry)
XX
XX Partial sequence of tumour suppressor gene U10.
XX
XX Can19; tumour suppressor gene; cancer; therapy; ss.
XX Homo sapiens.
XX WO9215602-A1.
XX 17-SEP-1992.
XX 28-FEB-1992; 92MO-US001624.
XX 28-FEB-1991; 91US-00662216.
XX (DAND ) DANA FARBER CANCER INST INC.
XX Sager R;
XX WPI; 1992-331663/40.
XX
XX Diagnosis and treatment of cancer - using candidate tumour suppressor
XX genes or the corresp. antibodies.
XX
XX Claim 30; Page 38-41; 54pp; English.
XX
XX An adaptation of the subtractive hybridization technique was used which
XX utilises a biotinylation-based subtraction procedure instead of
XX hydroxyapatite as previously used. In this procedure, a single strand
XX phagemid cDNA library from normal cell polyA+ mRNA is hybridized with
XX excess biotinylated tumor polyA+ mRNA, and the resulting double stranded
XX sequences are removed by binding to streptavidin. The remaining single-
XX stranded phagemid cDNAs are converted to double-stranded form and used to
XX transform bacterial host cells. The resulting subtracted cDNA library is
XX differentially screened with total cDNA from normal and tumor cells. This
XX method produced some 20 additional cloned cDNAs. Also found by this
```

CC method were several genes which, on the basis of the partial DNA  
 CC sequences appear to be novel sequences not previously entered into  
 CC GENBANK. The portion of the cDNAs so sequenced represents part of the  
 CC coding region and/or part of the 3' untranslated region of each cDNA (see  
 CC Q28749-58). (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-  
 CC MAR-2003 to correct PA field.)

XX  
 SQ Sequence 4328 BP; 1236 A; 970 C; 912 G; 1210 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 5,436-142 Length: 4328  
 Score: 3417.00 Matches: 648  
 Percent Similarity: 99.39% Conservative: 1  
 Best Local Similarity: 99.23% Mismatches: 3  
 Query Match: 64.80% Indels: 1  
 DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AAQ28758 (1-4328)

QY 355 GlnValysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374  
 DB 3 CAGGTGAAGACCAACCCCTGAGGAGAAGAGTCTTTGACCTTATTTCACATGACAGAACT 62  
 QY 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTyrMetSerValLeuGlnAsn 394  
 DB 63 TACCACCTTCAAGCTGAAGATGAACAGGAATGTCAATATATGGATGCTGTGCTGCAAAAT 122  
 QY 395 SerIysGluAlaLeuAsnAlaPheIysGlyAspAsnThrGlyGluAsnAsn 414  
 DB 123 AGCAAGAAGAAGCTTTAAACAATGCAATTTAAGGGGGATGACATATCTGGAGAAATTAAC 182  
 QY 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp 434  
 DB 183 ATGCTCCAAGAACTGACAAAGAGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATGAC 242  
 QY 435 ValCysCysAspCysGlyAlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeu 454  
 DB 243 GTCTGCTGTGACTGTGGGGGCCAGATCTTACATAGCTCTTCCACCACTGGGCATCTCTG 302  
 QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474  
 DB 303 ACCTGCTGAGTGTTCGGGAATCCACGAGAGCTGGGGGTTCATTATTCAGGATGCGAG 362  
 QY 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsn 494  
 DB 363 TCCCTGACCTTAGATGTGGAACTCTGAGCTGCTGCTCGCCAAAGAAATATTGGGAAT 422  
 QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValIysProAsn 514  
 DB 423 GCAGGCTTAAATGAGATCATGGAAATGTTGCCCTACCAGCTGAGGACTCAGTCAACCCCAAC 482  
 QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 534  
 DB 483 CCAGCGCAGCATGAATGCAAGAAGAGCTACATCACAGCCAAGTACATCGAGAGGAGA 542  
 QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVal 554  
 DB 543 TACGCAAGAGAAGACACGCGGATAAACGCGCGAAGCTTTCAGTCTTTTGGAGGCCGTC 602  
 QY 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574  
 DB 603 AAAACGAGAGATATTTTGGATGTCTCCNAGCTTATGCTGATGGTGTGATCTTACGGA 662  
 QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594  
 DB 663 AAAATCCCACTGGCCCAACGAGACATCAGCGGATGAACAGGGCCCTCCACTTGCAGTCAGA 722  
 QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerClyAsnLeu 614  
 DB 723 TCCGTGGATGAAACCTCTCTTCACATTTAGATCTTTTAGTTTCAAGACAGTGGGAACCTG 782  
 QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634

DB 783 GATAAACAGACAGGAGAAAGGAGCAGCACAGCCCTGCTCACTACTGCTGCACCAACAATGCC 842  
 QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGly 654  
 DB 843 GAGTGCCTCAAGTTGCTCTCTCGGGGAAGCCCTCCATCGAGATAGCAAAATGAGTCAGGA 902  
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674  
 DB 903 GAGACTCCGTGGACATTGCCAAGCGCTCAGACAGCACTGTGAGGAGCTGCTGACC 962  
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeu 694  
 DB 963 CAAGCCTTATCTGGAAGATTTAATTTCTCAGCTTCAAGTGAATGAATGCGGACTACTC 1022  
 QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714  
 DB 1023 CACGAAGACCTGGATGAAGTGAATGATGATGATGAGAAATTTGCAGCCCAAGTCCCAAC 1082  
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734  
 DB 1083 CGGCGGGAAGACCCGCCCATCAGCTTCTACAGCTGGGCTCCAACCAAGCTTCAGTCTAAC 1142  
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754  
 DB 1143 GCTGTATCTTTGGCCAGAGATGCTGCAAACTTGCCTGCAAGGACAGCAGAGGGCTTCATG 1202  
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774  
 DB 1203 CCCAGCATCTTGCAGAATGAGACTTACGAGGCCCTCTCGATGGCAGCCCACTCCCGCC 1262  
 QY 775 GlnProAlaLaProSerThrThrSerAla-ProProLeuProProArgAsnValGlyLys 794  
 DB 1263 CAGCTCGAGCCCCCAGCACCAACAGCGCCCCCCCCCTCTCTCCAGGAATGTGGCAA 1322  
 QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGly 814  
 DB 1323 AGTTTCAGACAGCCTCTCTGCTAACACCTGTGGAAGACAAACTCTGTAAGTGTGGACGG 1382  
 QY 814 YGlySerArgGlnArgSerSerSerAspProAlaValHisProProLeuProProLe 834  
 DB 1383 TGGAAAGCGCGCAGCGATCTTCTGTCAGATCCGCGAGCTGTCCATCCACCGCTGCCCTCT 1442  
 QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProValAlaLysThrPr 854  
 DB 1443 TCGGTGACATCTACCAATCCCCCTGACCCCGCCCGCCCGCTGTCGCAAGACGCC 1502  
 QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnI 874  
 DB 1503 CAGCGTAATGGAAGCCTTGAGCCAGCGAGCAAGCCTGCCCGCTGCGGATCTCACAGAT 1562  
 QY 874 eArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894  
 DB 1563 CAGGCCCCCACTCTGCCCCCAGCGCCCGCCCGCTCCCGCAGAAGAGCGCTGCGCC 1622  
 QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLe 914  
 DB 1623 AGGGGCTGACAACTTCAACCCCACTGACCAACAAAGGCCACCGAGAGACCTGTGGATCT 1682  
 QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934  
 DB 1683 CTCTGCAACGGAAGCTCTGGGTCTCTGTGCTCAATGCTATGGTCTCTGAGCCCCCTGCACC 1742  
 QY 934 oMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAs 954  
 DB 1743 CATCCCTTAGGAAGTCGAGGCAACCAAGTTGAAGCCTTAAGCGGGTGAAGAGCGCTCTATPA 1802  
 QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAs 974  
 DB 1803 CTGTGTGGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCATGGA 1862  
 QY 974 pGlyGluGluAspGlnGluTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGly 994  
 DB 1863 CGGGGAGGAGACCGAGGATGGTGGATTGGCCACATTTGATGATGATGATGATGATGATGATG 1922

QY 994 yAlaPheProValSerPheValHisPheAlaAlaasp 1006  
 DB 1923 CGCATTCGGGTGTCATTGTGCACTTTATCGCTGAC 1959

RESULT 3  
 AAV59106  
 ID AAV59106 standard; cDNA; 4595 BP.

XX AAV59106;  
 XX AC  
 XX 17-OCT-2003 (revised)  
 XX DT 20-NOV-1998 (first entry)

XX Zebrafish differentiation enhancing factor 3 gene.

XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
 KW C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
 KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
 KW nervous system.

XX Danio rerio.

XX Key Location/Qualifiers  
 XX FT 300..3011

XX FT /\*tag= a

XX FT /product= "DEF3 protein"

XX PN W09836065-A1.

XX PD 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US002724.

XX PR 14-FEB-1997; 97US-0038191P.

XX PA (DAND ) DANA FARBER CANCER INST INC.

XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;

XX WPI; 1998-467173/40.

XX P-PSDB; AAV59106.

XX New nucleic acid encoding differentiation enhancing factor - used  
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
 PT tumours and neurological disease.

XX Claim 7; Fig 15; 203pp; English.

XX The differentiation enhancing factors (DEF), comprise at least one each  
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
 CC are mediators of SH3-domain dependent signalling and may be involved in  
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
 CC endocytosis or adhesion, migration, proliferation and differentiation of  
 CC cells. Typical applications of DEF and agents that modulate interaction of  
 CC between the protein and it's ligand, or of nucleic acid expressing them,  
 CC are treatment of hyperplastic and neoplastic disease (a wide range of  
 CC solid tumours and leukaemias), including metastases; for in vitro  
 CC induction of differentiation of neural crest cells to neurons, glial  
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
 CC the nervous system (e.g. treatment of traumatic injury, stroke,  
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
 CC standardise OS field)

XX SQ Sequence 4595 BP; 1378 A; 1014 C; 1030 G; 1173 T; 0 U; 0 Other;

XX Alignment Scores:

Pred. No.: 2,14e-130 Length: 4595

Score: 3155.00 Matches: 621

Percent Similarity: 75.60% Conservative: 98

Best Local Similarity: 65.30% Mismatches: 174

Query Match: 59.83% Indels: 58  
 DB: 2 Gaps: 10  
 US-09-914-042-1 (1-1006) x AAV59106 (1-4595)

QY 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
 DB 300 ATGCTTGATCAGTCGACAGTCAGTGTGGATTATACCCATGAGGACTATATAAGCA 359  
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
 DB 360 CCGACACATCAGTGTCTGACCGCGCATGCTCCTGTCAGGAGTACAGTGGCGCTCTG 419  
 QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrIysMetIysSerValIysAla 60  
 DB 420 GAAGAGCGCTGGATCTGGACCGCAGTGTCTGCACAAATGAAGAGTCAAGGCC 479  
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
 DB 480 ATAAACAGCTCTGGTCAGACTCATGTAGAACGAGGAGGAGTACATCCAGGCCATAGAG 539  
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProaspLeuGlySerAlaPheLeuIys 100  
 DB 540 AGGTTTACGGATAACACTGTGTACAAGATGACCTCGAGATGTCCAATTACTTCTCACA 599  
 QY 101 PheSerValPheThrIysGluLeuThrAlaLeuPheIysAsnLeuIleGlnAsnMetAsn 120  
 DB 600 TTCGCTGTTTCAACAGGAGCTTACTGCTCTTTTCAAGAACTTGCTACAGAACATGAAT 659  
 QY 121 AsnIleIleSerPheProLeuAspSerIleLeuIysGlyAspLeuIysGlyValIysGly 140  
 DB 660 AACATCATCACTTTTCCACTAGACAGTCTGTAAGGGAGAGACCTCAAGGAGTCAAGGG 719  
 QY 141 AspLeuIysIysProPheAspIysAlaIleIysAspTyrGluThrIysIleThrIysIle 160  
 DB 720 GATTTGAAAAGCCATTGTAAGCATGGAGGATTATGAACCAAACTGAGCAGAAATT 779  
 QY 161 GluIysGluIysIysGluHisAlaIysLeuHisGlyMetIleIleThrGluIleSerGly 180  
 DB 780 GAGAAAGAAAAGCGAAGACATGCCAAACAGCAGCGTCTGTATCCGAACAGAGATCAGTGA 839  
 QY 181 AlaGluIleAlaGluGluMetGluIysGluArgPhePheGlnLeuGlnMetCysGlu 200  
 DB 840 GGAGAGATCGCAGAAAGAGATGGAGAAAGAGAGACGCTGTTTTCAGCTTCAGATGTGTAG 899  
 QY 201 TyrLeuLeuIysValAsnGluIleIysIleIysIysGlyValAspLeuLeuGlnAsnLeu 220  
 DB 900 TACCTCATTAAGTGAATGAATCAAGTCAAAAGGGGGTGGACCTCTTCAACACCTC 959  
 QY 221 IleIysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuIysAlaValGluSer 240  
 DB 960 ATCAAAATACTTTCATGCCAGTGCATTTCTTTTTCAGGATGGCTAAAGGTCTGTGACAAT 1019  
 QY 241 LeuIysProSerIleGluThrLeuSerThrAspLeuHisThrIleIysGlnAlaGlnAsp 260  
 DB 1020 CTGAACCTTTTCATGGAAAAGCTTCCACAGACTTAACCGCGAACAACAGACTCAAGT 1079  
 QY 261 GluGluArgGlnLeuIleGlnLeuArgAspIleLeuIysSerAlaLeuGlnValGlu 280  
 DB 1080 GCAGAAAGGAAGAACTGTGTCAGCTGAAGAAACTCTTAATCTGCTCTACAGTCTGAG 1139  
 QY 281 GlnIysGluAspSerGlnIleGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 DB 1140 TGTAAGGAGGATGCTCAGTCAAACGACAGACGCGCTACAGTCTTCCACAGTTCAGGGC 1199  
 QY 301 AsnIysGluHisGlyThrGluArgAsnGlySerLeuTyrIysIysSerAspGlyIleArg 320  
 DB 1200 AATAAGCTCACGCGAGCGCTCTGGATGCTCTCTCAAAACGACGAGGAGGACTGAGG 1259  
 QY 321 LysValTrpGlnIysArgIysCysSerValIysAsnGlyPheLeuThrIleSerHisGly 340  
 DB 1260 AAGTTTGGCAGAAAAGAAAGTGTCTGTGAAAAAATGATTTGTTGACTATTTCACATGGA 1319

```
Qy 341 ThrAlaAsnArgProAlaLysLeuLeuLeuThrCysGlnValLysThrAsnPro 360
|||
Db 1320 ACGCCCATCGACCGCAGCAAACTGAACCTCTTAACCTGCCAAGTGAAGCGTAACCCA 1379
|||
Qy 361 GluGluLysCysPheAspLeuLysSerHisAspArgThrTyrHisPheGlnAlaGlu 380
:::
Db 1380 GATGAGAAAAATGCTTTGATCTCATATCATGACAGAACGATCATCTCCAGACTGAG 1439
|||
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
|||
Db 1440 GATGAGGCGAGTGTGACGATGGGTCTTCTGTTCTCCAGAACAGTAAGAGAGCGCGCTG 1499
|||
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
|||
Db 1500 AACATGCTTTAAAGACATCAGNATGAGGAGGAGAAATAACATTTGTCGAGAGCTCACT 1559
|||
Qy 421 LysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
|||
Db 1560 AAGGCCATCGTGGGGGAAGTGAAGAAATGAGCGGCAATGACGTGTGCTGTGACTGTGGA 1619
|||
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
|||
Db 1620 GCTTCCAAATCAACATGCTCTCCACAAACCTGGGTGTGTTGATTGCAATGGAATGCTCT 1679
|||
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
|||
Db 1680 GGGATCCATCGGAAATGGGGTCCACTACTCTCCGAATACAGTCTCTGACACTGGACCTC 1739
|||
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
|||
Db 1740 TTAGGCACATCTGAACATATGCTTCTTAACAGTGTGGGAAATGACAGATTCAATGAAATC 1799
|||
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
|||
Db 1800 ATGGGAAGCAAACTGTCTCTCAGAG---ATCCCAAAACCCCTACCTTCTAGTGACATGCGAG 1856
|||
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
|||
Db 1857 GTACGAAAGACTTCATCAGCCCAATACACAGAGAGCGTTTCGCTCAGAGNAGATAT 1916
|||
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
|||
Db 1917 GCAGATAACGCGACTCGACTGCATGCTGTGATGATGAGTGAAGTCTCGGACATCTTC 1976
|||
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
|||
Db 1977 TCCCTGATCCAGGTCTATGCTGAAGGACTGGACCTGATGGAGACCAATTAATCAGCCTAAC 2036
|||
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
|||
Db 2037 CAACATGAACCGGCGAGACATCATCTATCTTGGCGTACGATGGTGGACGGAACCTCC 2096
|||
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
|||
Db 2097 CTCCTATTTGGAATCTTCTGTACAGAACAGTGGCAATTTAGACAAGCAGACGCCAAA 2156
|||
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
|||
Db 2157 GGAAGCACAGCGCTACATTTATGCTGCTGATGATNACAGTGAATGATGAAGCTGCTG 2216
|||
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
|||
Db 2217 CTGCGGGGGAAGCATCTGTACGATTAATGATGCTGGAGACTGCTCTGATTTG 2276
|||
Qy 661 AlatyArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
|||
Db 2277 GCGCAGCGTCTCAAACTCCAAATGCGAGGAGTGTGACTCAGCGCGCAGACGGGGAAG 2336
|||
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
|||
Db 2337 TTCAATGTCCATGTGCAATGATGACTGGCGTCTGCAATGAATGAGGATCTGGACCGAG 2396
|||
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
|||
```

```
Db 2397 AGCGAAGATGAGATGGAGGACAAG-----CCCATTCCTCAGCGCTGAGGAGCGTCCA 2450
|||
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
|||
Db 2451 ATAAGCTGTATAGTTCAGGCGAGTGGCCCCATGATGATGCCAACATCAGAGCGCTCTGGCTCGG 2510
|||
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
|||
Db 2511 GACGTGCCAATGGTGGTCAATATAAGCAGAGGCTTTTATTCGAGCATGATGATGAAC 2570
|||
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
|||
Db 2571 GAGACTTACGGCACCATGCTCGATCCCACTCTCCACCACTG-----GGTTTACCAGGA 2624
|||
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
|||
Db 2625 GTACCTGGCAATCCTCTTTTACCCCTCGGCCCTTGGGAAGGGA----- 2669
|||
Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
|||
Db 2670 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGTCA 2708
|||
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
|||
Db 2709 TGTTTCAGATCTCGCA-----AACCTCAAACTCTTGAAACAAATAAATACTCTGTGTAT 2759
|||
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
|||
Db 2760 GTTCTGCTCTCTGCTCTCTCCACCTCTCTGCA----- 2792
|||
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
|||
Db 2793 -----CCCAAGACACTCCACTCCAGATCCAAAGCCAGCTCTTCTCTCCAGCAGCC 2846
|||
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
|||
Db 2847 ACGGCTCCT-----CCTGCACCA-----TCCGCA 2870
|||
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
|||
Db 2871 CCGCTCTTATTCACCTCTCTCTCTCAGGCCAGCGCT----- 2909
|||
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
|||
Db 2910 -----GTAGTGGCCCTCGACCCAGTATGCCCATTCGCCACTTCGTCA 2945
|||
Qy 941 AlaThrLysLeuLysProLysArgValLysAla 951
|||
Db 2946 CTGACTGATGTCAAAAGTCTGCTGCTCTAAAGCC 2978
|||
RESULT 4
ADD71214
ID ADD71214 standard; cDNA; 3775 BP.
XX
AC ADD71214;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human intracellular signalling molecule INTS1G-51 cDNA SEQ ID NO:103.
KW human; intracellular signalling molecule; INTS1G; cytostatic;
KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
KW antiinflammatory; thymomimetic; gene therapy;
KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
KW immune disorder; inflammatory disorder; AIDS; allergy;
KW developmental disorder; Hypothyroidism; Cushing's syndrome;
KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.
OS Homo sapiens.
XX
```

PN WO2003039348-A2.  
 XX 15-MAY-2003.  
 XX 07-NOV-2002; 2002WO-US036151.  
 XX 09-NOV-2001; 2001US-0344472P.  
 PR 30-NOV-2001; 2001US-0334558P.  
 PR 14-DEC-2001; 2001US-0340296P.  
 PR 21-DEC-2001; 2001US-0343557P.  
 PR 18-JAN-2002; 2002US-0350420P.  
 PR 25-JAN-2002; 2002US-0351927P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Kable AE, Swarnakar A, Gervad AE, Hafalia AJA, Duggan BM;  
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;  
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;  
 PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;  
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare S, Becha SD;  
 PI Lee SY, Tran UK, Elliott VS, Sprague WW, Tang YT, Zebajadian Y;  
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;  
 PI Zheng W;  
 XX WPI; 2003-441441/41.  
 DR P-PSDB; ADD71162.  
 XX New human intracellular signaling molecules (INTSIG)), useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or  
 PT infections.  
 XX Claim 5; SEQ ID NO 103; 363pp; English.  
 XX The present invention describes human intracellular signalling molecules  
 CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins  
 CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,  
 CC antiinflammatory and thymomimetic activities, and can be used in gene  
 CC therapy. The INTSIG polynucleotides and proteins can be used in the  
 CC diagnosis, treatment and prevention of diseases or conditions associated  
 CC with the decreased expression or overexpression of INTSIG, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),  
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),  
 CC immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.  
 CC hypothyroidism, Cushing's syndrome), reproductive and vesicle-trafficking  
 CC disorders, or infections. They are also useful in assessing the effects  
 CC of exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of INTSIG. The INTSIG or its fragments are useful in screening  
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
 CC or in altering the expression of the target polynucleotide and compounds  
 CC that specifically bind to or modulate the activity of the polypeptide.  
 CC The present sequence encodes human INTSIG-51 from the present invention.  
 XX  
 SQ Sequence 3775 BP; 1123 A; 989 C; 873 G; 790 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 6 47e-125 Length: 3775  
 Score: 3029.00 Matches: 619  
 Percent Similarity: 69.78% Conservative: 129  
 Best Local Similarity: 57.74% Mismatches: 238  
 Query Match: 57.44% Indels: 86  
 DB: 10 Gaps: 19  
 US-09-914-042-1 (1-1006) x ADD71214 (1-3775)  
 QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
 DB 318 ATGCCGGACCATCTCTGCTCGGAGTTTCATCGCGGACACCCGAGGACTACAACTCG 377  
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 40  
 DB 378 CCCACCACTGTCAGCTTACCAACCGGGTGCACAACTGCAGGAACACCGCTCACTGCTG 437

QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
 DB GAGGAGGCTCTAGACCAAGATAGAACGCCCTTCAGAAAGTGAAGAGCTGTGTAAAGCA 497  
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
 DB ATATATAATCTGGTCAAGATCATGTACAAAATGAAGAAAACATATGCAACATCTTGTAT 557  
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
 DB AAGTTGGAGTAAATTTTAAAGTCGAGCAACCCCGACCTTGGCACCCTGCTTGTCAAG 617  
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120  
 DB TTTTCTACTCTTACAAAGAACTGTCCACACTGCTGAAAAAATCTGCTCCAGGTTTGAGC 677  
 QY 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
 DB CACATGTGATCTTCCACTTGGATCTTTGTTTAAAGGAGACCTTAAGGGAGTCAAGGA 737  
 QY 141 AspLeuLysLysProPheAspLysAlaLysAspTyrGluThrLysLysLysLys 160  
 DB GATCTCAAGAACCCATTTGACAAAGCCTGGAAAGATTATGACAAAGTTTACAAATTT 797  
 QY 161 GluLysGluLysLysGluHisAlaLysLysLysGlyMetIleArgThrGluLysSerGly 180  
 DB GAGAAAGAGAAAGAGAGAGACGACCAAAACCAACATGGGATGATCCGCACAGAGATAACAG 857  
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnGlnGlnMetCysGlu 200  
 DB GCTGAGATTGGGAGAAATGGAGAAAGGAGGCGCTCTTTTCACTCCAAATGTGTGA 917  
 QY 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysLysLysLysLysLysLysLys 220  
 DB TATCTCATTAAGTTAATGAATCAAGACCAAAAGGGTGTGGATCTCTCGCAGATCTT 977  
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLysLysAlaValGluSer 240  
 DB ATAAAGTATTACCATGCACAGTGCATTTCTTCAAGATGGCTTCAAAACAGCTGATAAG 1037  
 QY 241 LeuLysProSerIleGluThrLysSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 DB TTGAACAGTAGTACATTTGAAAAAATGGCTCTGCTGATTTATATATATATAAACAAGACCCAGGAT 1097  
 QY 261 GluGluArgArgGlnLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 280  
 DB GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1157  
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 DB CAGAAAGAAAGATTTCTCAGAGCGGCAAGGA--GGATACAGCATGCATCAGCTCCAGGCG 1214  
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLysArg 320  
 DB AATAGGAATATGGCAGTGAAGAAAGGAGTACCTGTCTTAAAGAAAGTGAAGCGGATCCGG 1274  
 QY 321 LysValTyrGlnLysArgLysCysSerValLysValAsnGlyPheLeuThrLysHisGly 340  
 DB AAGATATGCGAG 1334  
 QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
 DB ACATCTAACAGGCAACAGCCCAAGTTGAACCTTCTCACCTGCCAAGTAAACCTTAATGCC 1394  
 QY 361 GluLysLysLysCysPheAspLeuLysSerHisAspArgThrThrHisPheGlnAlaGlu 380  
 DB GAAGACAAAAAATCTTTGACCTGATATACATATAGAACATATACATCTTTTCAAGCAGAA 1454  
 QY 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluLysAlaLeu 400  
 DB GATGAGCAGGATTTATGTAGCATGGATATCATGATTGTGACAAATAGCAAAAGAGAGGCCCTTA 1514

```
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1515 ACCATGGCCCTTCGTTGGAGACAGAGTCGGGAGAGAACACGCTG---GAAGACCTGACA 1571
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1572 AAAGCCATTATTAGGATGTCACGCGCTCCACGGGATGACATTTGCTGCGATTGTGC 1631
Qy 441 AlaProAspProThrTrrPLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1632 TCATCAGAACCCACCTGCTCTTCAACCAACTGGGTATTTTGACCTGTATAGAATGTTCT 1691
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1692 GGCATCCATAGGAAATGGGGTTTCATATTCTCGCATTCAGTCTTTTGGAACTAGACAAA 1751
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1752 TTAGAACTTCTGAACCTTCTGCTGCGCAAGATGTAGGAAACAATAGTTTAAATGATATT 1811
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1812 ATGGAAGCAAAATTTACCCAGC---CCCTCACCAAAACCCACCCCTTCAAGTGATAGTACT 1868
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHis 540
Db 1869 GTACGAAAGAATATATCAGTCGCAAGATGTAGATCATAGTGTTCGAAAGAACCTGT 1928
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1929 TCAACTTCATCAGCTAAATAAATGAATTGCTTGAGGCCATCAATCCAGGGATTTACTT 1988
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1989 GCACATAATCAAGTCTATGCGAGGGGTAGAGCTAATGGAACCACTG---CTGGAACCT 2045
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2046 GGGCAGGAGCTTGGGAGACAGCCCTTACCTTGGCCGTCCGAACTGCGAGATCAGACATCT 2105
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2106 CTCCTATTGGTTGACTCTCTGTACAAACCTGTGGGAACCTGGATAAGCAGCGCCCTG 2165
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2166 GGAACACACAGCTTCTACACTACTGTATGTATGATACAGTAAACCTGAGTGTTTGAAAGCTTTG 2225
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2226 CTCAGGACCAAGCCACTGTGGATATAGTTAACAGGCTGGAGAACTGCCCCTAGACATA 2285
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2286 GCAAAAGACATAAAAGCTACCCAGTGTGAAGATCTGCTTCCAGGCTAAATCTGGAAG 2345
Qy 681 PheAsnSerHisValHisValGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2346 TTCAATCCACAGCTCCAGCTAGAATATAGTGTGAATCTTCGACAGGAGGATAGATGAG 2405
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2406 ACCGATGATGATCTGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCCAGACT 2465
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerIleAlaArg 740
Db 2466 CAGAGCTTCTGCCACTCTCTCCAGCATCTCCCCCAGGAC-----AAGCTGGCACTG 2516
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2517 CCAGGATTTCAGCACTTCAAGGGACAAACAGCGGCTC----- 2552
Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla 777
```

```
Db 2553 ---TCCTATGGAGCTTCCACACCAGATCTTCGTTTCCACACGACAGACTCCGCCACA 2609
Qy 778 AlaProSerThrThrSerAlaProLeuProProArgAsnValGlyLysValGlnThr 797
Db 2610 TCACCA---ACCACGAGGCTCCCTCTCTAGGAACGCGGGAAGGTAACGAT 2666
Qy 798 AlaSerSerAlaAsnThrLeuTrrPLeuThrAsnSerValSerValAspGlyLysArg 817
Db 2667 GGGGGTCCATCTCTTCAAGTAAGACTACAAC-----AAGTTTGAGGGACTATCCAG 2720
Qy 818 GlnArgSerSerSerAspPro-----ProAlaValHisProProLeuPro 832
Db 2721 CAGTCGAGCACCAGTCTTGCAAGACTGCCCTTGGCCCAAGAGTCTTCTCTAAACTACT 2780
Qy 833 -----ProLeuArgValThrSerThrAsnProLeu----- 842
Db 2781 CAGAAAGTGGCACTTAAGGAAAAACAGATCATCTCTCCCTAGACAAAGCCACCATCCGCC 2840
Qy 843 -----ThrProThrProProProPro----- 849
Db 2841 GAAATCTTTCAGAAATCATCACAGTTGGCAGAGTTGCCACAAAGCCACCACTGGAGAC 2900
Qy 850 -----ValAlaLysThrProSerValMetGluAlaLeuSerGln 862
Db 2901 CTGCCCCCAAGCCACAGAACTGGCCCCCAAGCCCAATTTGGAGATTTGCCGCCCTAAG 2960
Qy 863 ProSerLys---ProAlaProProGlyIleSerGlnIleArgProProPro----- 878
Db 2961 CCAGGAGAACTGCCCCCAACCAACACAGCTGGGGGACCTGCCCAACCAACCCCAACTCTCA 3020
Qy 879 ---LeuProProGlnProProSerArg---LeuProGlnLys----- 890
Db 3021 GACTTACTCTCCAAACACACAGATGAAGACCTGCCCCCCCAACCAACCAAGCTGGAGACCTG 3080
Qy 891 -----LysProAlaProGlyThr 896
Db 3081 CTAGCAAAATCCCAGACTGGAGATGTCTCACCCAAGGCTCAGCAACCTCTCTGAGGTGACA 3140
Qy 897 AspLysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeu 914
Db 3141 CTGAAGTCACACACCTTGGATCTATCCCAAAATGTGCAGTCCAGAGACGCCATCCAAAAG 3200
Qy 915 SerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPro 934
Db 3201 CAAGCATCTGAAGACTCCAAGACCTCAGGCTACTCTGCCA---GAGACCCCGTACCA 3257
Qy 935 MetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsn 954
Db 3258 CTGCCCAGAAAAATCAATACGGGAAAAATAAAGTGAGGGAGTGAAGACCATTTATGAC 3317
Qy 955 CysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAsp 974
Db 3318 TGCCAGGACAGACACAGTACGAGCTCACATTCATCAGGGGAGAAAGTATTATCGTCACA 3377
Qy 975 GlyGluGluAspGlnGluTrrPileGlyHisIleAspGlyAspProGlyArgLysGly 994
Db 3378 GGGGAAGAGGACCAAGAGTGTGGATTGGCCCAATCGAAGGACAGCTTGAAGAGAGGGG 3437
Qy 995 AlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3438 GTCTTCCAGTGTCTTTGTTTCATATCTCTGTCTGAC 3473

RESULT 5
ADD71207
ID ADD71207 standard; cDNA; 3946 BP.
XX
AC ADD71207;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human intracellular signalling molecule INTSIG-44 cDNA SEQ ID NO:96.
XX
```

KW human; intracellular signalling molecule; INTS1G; cytoskeletal;  
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;  
 KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;  
 KW antiinflammatory; thyroidemic; gene therapy;  
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;  
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;  
 KW immune disorder; inflammatory disorder; AIDS; allergy;  
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;  
 KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.

XX Homo sapiens.  
 XX WO2003039348-A2.  
 XX 15-MAY-2003.  
 XX 07-NOV-2002; 2002WO-US036151.  
 XX 09-NOV-2001; 2001US-0344472P.  
 XX 30-NOV-2001; 2001US-0334558P.  
 XX 14-DEC-2001; 2001US-0340296P.  
 XX 21-DEC-2001; 2001US-0343557P.  
 XX 18-JAN-2002; 2002US-0350420P.  
 XX 25-JAN-2002; 2002US-0351927P.  
 XX (INCYT) INCYTE GENOMICS INC.  
 XX Kable AB, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;  
 XX Warren BA, Emerling BM, Leon CH, Nguyen DB, Lindquist EA;  
 XX Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li UX;  
 XX Marquis JP, Gierzen KJ, Baugman MR, Borowsky ML, Yao MG, Chawla NK;  
 XX Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;  
 XX Lee SY, Tran UK, Elliott VS, Sprague WW, Tang YT, Zebajadian Y;  
 XX Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;  
 XX Zheng W;  
 XX WPI; 2003-441441/41.  
 XX P-PSDB; ADD71155.  
 XX New human intracellular signaling molecules (INTS1G), useful for  
 XX diagnosing, treating and preventing diseases or conditions associated  
 XX with the aberrant INTS1G expression e.g. cancer, diabetes, epilepsy, or  
 XX infections.  
 XX Claim 5; SEQ ID NO 96; 363pp; English.  
 XX The present invention describes human intracellular signalling molecules  
 XX designated INTS1G-1 to INTS1G-52. The INTS1G polynucleotides and proteins  
 XX have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,  
 XX nootropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,  
 XX antiinflammatory and thymomimetic activities, and can be used in gene  
 XX therapy. The INTS1G polynucleotides and proteins can be used in the  
 XX diagnosis, treatment and prevention of diseases or conditions associated  
 XX with the decreased expression or overexpression of INTS1G, such as cell  
 XX proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),  
 XX neurological (e.g. epilepsy, Huntington's disease, stroke),  
 XX immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.  
 XX hypothyroidism, Cushing's syndrome), reproductive and vesicle-trafficking  
 XX disorders, or infections. They are also useful in assessing the effects  
 XX of exogenous compounds on the expression of nucleic acid and amino acid  
 XX sequences of INTS1G. The INTS1G or its fragments are useful in screening  
 XX compounds for effectiveness as agonist or antagonist of the polypeptides,  
 XX or in altering the expression of the target polynucleotide and compounds  
 XX that specifically bind to or modulate the activity of the polypeptide.  
 XX The present sequence encodes human INTS1G-44 from the present invention.

XX SQ Sequence 3946 BP; 1165 A; 1064 C; 902 G; 815 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1,078-124 Length: 3946  
 Score: 3024.50 Matches: 627  
 Percent Similarity: 66.84% Conservative: 127  
 Best Local Similarity: 55.59% Mismatches: 233

Query Match: 57.36% Indels: 141  
 DB: 10 Gaps: 21  
 US-09-914-042-1 (1-1006) x ADD71207 (1-3946)  
 QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20  
 DB 318 ATGCCGACACAGATCTCTGTCGAGTTCATCGCGAGACACCGAGACTCAACTCG 377  
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaIle 40  
 DB 378 CCCACACGTCAGCTTCCACACGCGGTGCACACTGCAGGAAACACCGTCAGCTGCTG 437  
 QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60  
 DB 438 GAGGAGGCTTAGACCAAGATAGAACAGCCCTTTCAGAAAGTGAAGAAGTCTCTAAAGCA 497  
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluTyrThrGlnAlaLeuGlu 80  
 DB 498 ATATATAATTCGTGTCAGATCATGTACAAATGAAGAAACTATGCACAGTCTTGTAT 557  
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
 DB 558 AAGTTGGAGTATTTTAAAGTCAGACAAACCCCGACCTTGGCACCCTTGTGCAAG 617  
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120  
 DB 618 TTTTCTACTTTACAAAGGAACTGTCACACTGCTGAAATAATCTGCTCCAGGTTTGAGC 677  
 QY 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValIysGly 140  
 DB 678 CACATGTCGCTTCCACCTTGGATCTTTTGTAAAGGAGAGACTAAAGGAGTCAAGGA 737  
 QY 141 AspLeuLysPheProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
 DB 738 GATCTCAAGAAGCCATTTGACAAAGCTGGAAGATTTATGACAAAGTTTTCAAAAT 797  
 QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180  
 DB 798 GAGAAAGAGAAAGAGAGAGACGACCAACCAACATGGGATGATCCGCACAGAGATACAGGA 857  
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGluLeuGlnMetCysGlu 200  
 DB 858 GCTGAGATTGGGAGAAATGGAGAAAGGAGGCGCTCTTTTCAGCTCCAAATGTGTGA 917  
 QY 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysLysLysLysLysLysLysLys 220  
 DB 918 TATCTCATTAAGTTTAAAGAAATCAAGACCAAAAGGGTGTGGATCTGCTGAGAAATCTT 977  
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLysLysAlaValGluSer 240  
 DB 978 ATAAAGTATTACCATGCACAGTGCATTTCTTCAAGATGGCTTCANAAACAGCTGATAG 1037  
 QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 DB 1038 TTGAAACAGTACATTTGAAATACTGCTGCTGATTTATATATATATATATATATATAT 1097  
 QY 261 GluGluArgArgGlnLeuIleGlnLeuArgPheIleLeuLysSerAlaLeuGlnValGlu 280  
 DB 1098 GAAGAAAGAAAGAAAGTCTGACCTCCGAGACTTAAATAATCTCTCTTCACTGAT 1157  
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 DB 1158 CAGAAAGAAAGATTTCTCAGACGCGGCAAGGA---GGATACAGCATCATCAGCTCCAGGC 1214  
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
 DB 1215 AATAAGGAATATGCGAGTGAAGAGGGGTACCTGCTTAAAGAAAGTGAAGCGGATCCGG 1274  
 QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
 DB 1275 AAGTATGCGAGAGAGAGAGTGTTCAGTCAAGATGGGATTTCTGACCATCTCATCATGCC 1334



Qy	341	ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnVallysThrAsnPro	360
Db	1335	ACATCTAACAGGCAACCCAGCAAGTTGAACTTCTCACCTGCCCAAGTAAACCTAATGCG	1394
Qy	361	GluGluLysCysPheAspLeuLeuSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1395	GAGAGCAAAAATATCTTTTGACCTGATATACATAATAGAACATATCACCTTTCAGGCAGAA	1454
Qy	381	AspGluGlnGluCysGlnLeuLeuTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
Db	1455	GATGAGCAGGATATATGTAGCATGGATATCAGTATTGCAATATAGCAAGNAGAGGCCCTA	1514
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnLeuValGlnGluLeuThr	420
Db	1515	ACCATGCCCTTCGTGGAGAGCAGAGTCGCGGAGAGAACAGACCTG--GAAGACCTTGACA	1571
Qy	421	LysGluLeuLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1572	AAAGCCATATTTAGGAGATGTCACGCGCTCCCAAGGGAATGACATTTGCTGCGAATGTGGC	1631
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyLeuLeuThrCysLeuGluCysSer	460
Db	1632	TCATCAGAACCCACCTGGCTTTCAACCACTTTGGTATTTTGGACCTGTATAGAAATGTCT	1691
Qy	461	GlyLeuHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1692	GGCATCCATAGGGAATGGGGTTCATATTTCTCGCATTCAGTCTTTGGAACTAGACAAA	1751
Qy	481	LeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluLeu	500
Db	1752	TTAGGAATCTTCTGAACCTTCTGCTGGCCAGAAATGTAGGAAACAATATAGTTTAAATGATAT	1811
Qy	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
Db	1812	ATGGNACAAATTTACCACG--CCCTCACAAACCACCCCTTCAAGTGATATGACT	1868
Qy	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHis	540
Db	1869	GTACGAAAAGAAATATACACTGCAAAAGTATGTAGATCATAGTTTTCAGGAAGACCTGT	1928
Qy	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVallysThrArgAspIlePhe	560
Db	1929	TCAACTTCATCAGCTAAACTAAATAGAAATTCCTTGAGGCCATCAAAATCCAGGGATTTACTT	1988
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	1989	GCACATAATCAAGTCTATGACAGAGGGGTAGAGCTAATGGAAACCACCTG---CTGGAACT	2045
Qy	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
Db	2046	GGGCAGGAGCTTGGGGAGACAGCCCTTCACCTTGCCTGCCGATGCAGATCAGACATCT	2105
Qy	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620
Db	2106	CTCCATTTGGTTGATCTTCCTTGTCACAAAACCTGTGGGAACCTGGATAGCAGACGGCCCTG	2165
Qy	621	GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
Db	2166	GGAAACACAGTTCTACACTACTGTAGTATGTACAGTAAACCTGAGTGTTTGAAGCTTTTG	2225
Qy	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
Db	2226	CTCAGGAGCAGCCCACTGTGGATATAGTTAAACCAGCGTCGGAGAAACCTGCCCTAGACATA	2285
Qy	661	AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg	680
Db	2286	GCAAAAGAGACTAAAGACTTACCAGCTGTGAAGATCTGCTTTCCAGGGCTAAATCTGGAAAG	2345
Qy	681	PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu	700
Db	2346	TTCAATCCACACGCTCCACGTAGAATATAGTGGAAATCTTCCACAGGAGGAGATAGATGAG	2405
Qy	701	SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro	720

Db	2406	GATGATCTGGATGATCAAAACCAAGCCCTTATCAAGAAAGAGCGCTCACC	CCGACCT	2465
Qy	721	IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg	740	
Db	2466	CAGAGCTTCGCACCTCTCCAGCATCTCCCCCAGGAC-----AAGCTGGCAGCTG	2516	
Qy	741	AspAlaAlaAsnLeuAlaIysGluIysGlnArgAlaPheMetProSerIleLeuGlnAsn	760	
Db	2517	CCAGGATTACGACTCCAGGGAGCAACAGCGGCTC-----	2552	
Qy	761	GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla	777	
Db	2553	---TCCTATGGAGCCTTACCACACAGATCTTCGTTTCCACAGCAGACTCGCCCA	2609	
Qy	778	AlaProSerThrThrSerAlaProLeuProArgAsnValGlyLys-----	794	
Db	2610	TCACCA---ACCACGAGGCTCCCTCTGCTCTCTAGGAACGCCGGGAAAGGTCCAACT	2666	
Qy	795	-----ValGlnThrAlaSerSerAlaAsnThrLeuTyr	805	
Db	2667	GGCCACCTTCAACACTCCCTCTAAGCACCAGACTCTAGTGGCAGCTCCACCTTATCC	2726	
Qy	806	LysThrAsnSerValSerValAspGlyGlySerArgGlnArgSerSerSerAspProPro	825	
Db	2727	AAGAAGAGGCTCTCTCCCCACACCCCGGA---CACAAAGAGAACCTTATCCGACCTCC	2783	
Qy	826	AlaValHisProLeuProLeuLeuArg-----	835	
Db	2784	AGCCACTACTCATGGGCCCCCAACAAAGGCGCAGTCTCTTGGGGTACGATGGGGT	2843	
Qy	835	-----	835	
Db	2844	CCATCTCTTCAAGTAAGACTACAAACAAGTTTGGGGACTATCCAGCAGTCGAGCACC	2903	
Qy	836	-----ValThrSerThrAsnPro---LeuThrProThrProProProValAla	851	
Db	2904	AGTTCTGCAAGAGCTGCGCTTGGCCCAAGAGTTCCTTAACTACTCAGAAAGTGGCA	2963	
Qy	852	-----LysThr-----ProSerValMetGlu	858	
Db	2964	CTAAGGAAACAGATCATCTCTCCCTAGACAAAGCCACCCTTGGAGACCTGCCCCCAAG	3083	
Qy	859	AlaLeuSerGln-----ProSerLysProAlaPro-----	868	
Db	3024	AAATCATCACAGTTGGCAGAGTTGCCACAAAGCCACCCTTGGAGACCTGCCCCCAAG	3083	
Qy	869	-----ProGlyIleSerGlnIleArgProProPro-----	878	
Db	3084	CCCACAGAACTGGCCCCCAAGCCCAATTTGGAGATTTGCGCGCTAAGCCAGGAGAACTG	3143	
Qy	879	-----LeuProGlnProPro---SerArgLeuProGln	889	
Db	3144	CCCCCAACACACAGTGGGGGACCTGCCACCCAAACCCCACTCTCAGACTTACTCTCC	3203	
Qy	890	Lys-----	890	
Db	3204	AAACCACAGATGAAGGACCTGCCCCCAACACACAGCTGGGAGACCTGTAGCAAAATCC	3263	
Qy	891	-----LysProAlaProGlyThrAspLysSerThr	900	
Db	3264	CAGACTGGAGATGTCTCACCACAAAGGCTCAGCAACCCCTCTGAGGTCACTGAAGTCA	3323	
Qy	901	Pro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGlu	918	
Db	3324	CCATTGATCTATCCCCAAATGTGCAGTCAGAGACGCCATCCAAAGCAAGCATCTGAA	3383	
Qy	919	AlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLys	938	
Db	3384	GACTCCAAAGCCTCAGCGCTACTCTGCCA---GAGACGCCCGTACCCTGCCCAGAAAA	3440	
Qy	939	SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAsp	958	



Db 3441 ATCAATACGGGGAAAAATAAAGTAGGCGGAGTAGGAGAACCAATTATGACTGCCAGGCAGAC 3500

Qy 959 AsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAsp 978  
 |||||  
 Db 3501 AACGATGACGAGCTCACATTCATCGAGGAGAGAGTATTATCGTCACAGGGGAGAGGAC 3560  
 |||||  
 Qy 979 GlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProVal 998  
 |||||  
 Db 3561 CAGGAGTGTGGATTGGCCACATCGAAGCAGACGCCCTGAAGGAGAGGGGGTCTTTCCAGTG 3620  
 |||||

Qy 999 SerPheValHisPheIleAlaAsp 1006  
 |||||  
 Db 3621 TCCITTTGTTATATCTGTCTGAC 3644  
 |||||

RESULT 6  
 ID AAV59103  
 XX AAV59103 standard; cDNA; 5330 BP.  
 XX AAV59103;  
 XX 20-NOV-1998 (first entry)  
 XX Bovine differentiation enhancing factor 1 gene.  
 XX ss; bovine; differentiation enhancing factor; ankyrin repeat; C2 domain;  
 KW SH3 consensus binding sequence; pleckstrin homology domain; adipogenesis;  
 KW neurogenesis; hyperplastic disease; neoplastic disease; nervous system.  
 XX Bos sp.  
 XX Key Location/Qualifiers  
 FH CDS 209..3598  
 FT /\*tag= a  
 FT /product= "DEF1 protein"  
 XX W09836065-A1.  
 XX 20-AUG-1998.  
 XX 13-FEB-1998; 98WO-US002724.  
 XX 14-FEB-1997; 97US-0038191P.  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;  
 WPI; 1998-467173/40.  
 P-PSDB; AAW77286.  
 New nucleic acid encoding differentiation enhancing factor - used  
 particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
 tumours and neurological disease.  
 PS Claim 10; Fig 2; 203pp; English.

XX The differentiation enhancing factors (DEF), comprise at least one each  
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
 CC are mediators of SH3-domain dependent signalling and may be involved in  
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
 CC endocytosis or adhesion, migration, proliferation and differentiation of  
 CC cells. Typical applications of DEF and agents that modulate interaction  
 CC between the protein and it's ligand, or of nucleic acid expressing them,  
 CC are treatment of hyperplastic and neoplastic disease (a wide range of  
 CC solid tumours and leukaemias), including metastases; for in vitro  
 CC induction of differentiation of neural crest cells to neurons, glial  
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
 CC the nervous system (e.g. treatment of traumatic injury, stroke,  
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
 CC sclerosis, multiple sclerosis etc.)

XX Sequence 5330 BP; 1416 A; 1502 C; 1276 G; 1136 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,396-124 Length: 5330  
 Score: 3013.50 Matches: 627  
 Percent Similarity: 66.67% Conservative: 129  
 Best Local Similarity: 55.29% Mismatches: 225  
 Query Match: 57.15% Indels: 153  
 DB: 2 Gaps: 23

US-09-914-042-1 (1-1006) x AAV59103 (1-5330)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
 |||||  
 Db 269 ATGCGGACCGAGATCTCGGTCTCGAGTTATCGCCGAGACCCAGGAGACTTACAACTCG 328  
 |||||

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40  
 |||||  
 Db 329 CCCACCAGCTCCAGCTTCACCTACGCGGCTGCACAACTGCAGGAACACCGTCACGCTGCTG 388  
 |||||

Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60  
 |||||  
 Db 389 GAGGAGGCTCTAGACCAAGATAGAACAGCCTTACAGAAAGTTAAGAAGTCTGTAAGAAGCA 448  
 |||||

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80  
 |||||  
 Db 449 ATATACAATTCCGGTCAAGACCATGTACAAATGAAGAAACTATGCGCAAGTCTTCTGAT 508  
 |||||

Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100  
 |||||  
 Db 509 AAGTTGGAGTAATTTTTAAGTCGACAAACCCAGATCTTGGCACCCTTTTGTCAAG 568  
 |||||

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAsn 120  
 |||||  
 Db 569 TTTTCTACGCTTACAAAGGAACTGTCCACACTGTCTGTAATAATCTCTCCAGGCGCTGAGC 628  
 |||||

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140  
 |||||  
 Db 629 CACAATGTGATCTTTCACCTTGGATTCTTGTGTAAGGAGAGACCTCAAGGGAGTCAAGGC 688  
 |||||

Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160  
 |||||  
 Db 689 GATCTCAAGAAACCAATTGACAAAGCTTGGAAAGATTATGACGAGAAAGTTTACCAAAATT 748  
 |||||

Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180  
 |||||  
 Db 749 GAGAAGGAGAGAGGAGGAGCAGCCCAAGCAGCAGCGGATGATCCGACGAGATCACCGGC 808  
 |||||

Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200  
 |||||  
 Db 809 CCCGAGATCGCGGAGGAAATGGAAGAGGCGGCGCTCTTCCAGCTCCAGATGTGCGAG 868  
 |||||

Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220  
 |||||  
 Db 869 TATCTCATTAAGTTAATGAATCAAGACCAAAAGGGGTGGATCTGCTGCGAAGACCTG 928  
 |||||

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240  
 |||||  
 Db 929 ATAAAGTATTATACGACAGTCGCAATTTCTTTCAAGATGGTTTCAAAACAGCTGATAAA 988  
 |||||

Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260  
 |||||  
 Db 989 TTGAACAGTAGTACATTTGAAAGCTGCTGCTGATTGTTATATATATATCAACAGACCCAGGAC 1048  
 |||||

Qy 261 GluGluArgArgGlnLeuLeuGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280  
 |||||  
 Db 1049 GAAGAAAAGAAACAGCTACCGCAGCTCCGAGACCTAATAAAGTCTCTCCTCCACTCCAT 1108  
 |||||

Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297  
 |||||  
 Db 1109 CAGAAGGAGTCTAGGAGAGATTTCCAGAGCGGCGAGGGA---GGCTACAGCATGCACCAG 1165  
 |||||

Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
 |||||

Db 1166 CTGCAGGCAACAAGGAATACGGCAGCAGAGGAAGGGCTACCTCTCGAAGAAGAGTGAC 1225  
Qy GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
Db 1226 GGGATCCGGAAAGTGTGGCAGAGAAGAGTGTCTCCGTCAAGAACGGGATCTTGACCATC 1285  
Qy SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
Db 1286 TCCACAGCCAGTCCACACAGACAGCAGCAGCAAGCTGAACCTTCTCACATTGCCAGGTGAAG 1345  
Qy ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377  
Db 1346 CCGAATGCCGAGGACAAGAAGTCTTTTGACCTGATATCACATAACAGCAGCGTATCACITTT 1405  
Qy GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
Db 1406 CAGGCCGAAGATGACGAGGATATGTAGCTGGATCTCGGTCTGCACAAACAGCAAGAG 1465  
Qy GluAlaLeuAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGln 417  
Db 1466 GAGGCCCTCACCATGGCTCTTCGGGGGAACAGAGTGTGGGAGCAGCGCTG---GAG 1522  
Qy GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437  
Db 1523 GAGCTGACGAAGGCCATCATGAGGACGTGCAGCGGCTCCCGGGCAACGACGTCTGTCTGC 1582  
Qy AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGluIleLeuThrCysIle 457  
Db 1583 GACTCGCGCTCGGCAGAACCCACCTGGCTGTCCACCAACTTGGGCATCTTGACCTGTATA 1642  
Qy GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477  
Db 1643 GAATGTTCCGGCATCCATAGAAATGGGGTTCATATTTCTCGCATCCAGTCTTTGGAA 1702  
Qy LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497  
Db 1703 CTAGACAAATAGGAACCTCTGAACTCTTGCTGGCCAAAGATGTAGGAACAATAGTTT 1762  
Qy AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517  
Db 1763 AATGATATTATGGAAGCAAAATTTACCAGT---CCCTCACCACAAACCCACCCCTTCAAGT 1819  
Qy AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537  
Db 1820 GATATGACTGTACGGAAGAATATATCACTGCAAAAGTATGTAGATCATAGGTTTTCACGG 1879  
Qy LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557  
Db 1880 AAGACCTGTTTCATCGTCATCAGCTAAACTGAACGAATTTGCTTGAGGCCATCAAAATCCAGG 1939  
Qy AspllePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577  
Db 1940 GATTTACTTGCACTAAATTCAGTCTATGACAGGGGGTGGAGCTAATGGAACCCGCTG--- 1996  
Qy LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597  
Db 1997 CTGNAACCCGACAGAGCTTGGGAGACAGCCCTTCATCTTTCAGTCCGAAACCCACAG 2056  
Qy ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617  
Db 2057 CAGACATCTCTCCATTTGGTGACTTCTTGTGTACAAAACGTGGGAACTTAGATAAGCAG 2116  
Qy ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuLeuThrAspAsnAlaGluCysLeu 637  
Db 2117 ACGGCCCTGGGGAAACACGCCCTTGCACTACTGTAGTATGTACAGTAAACACAGAGTGTG 2176  
Qy LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657  
Db 2177 AAGTGTCTGCTCAGAGCAGACCCCACTGTGGACGTCTGTTAATCAGGCTGGAGAGACGCC 2236  
Qy LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677  
Db 2237 CTGGACATAGCAAGAGACTGAAGCCACTCAGTGTGAAGACCTGCTTTTCCCAAGCTAAA 2296

Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697  
Db 2297 TCTGGAAGTTCAATCCACAGCTCCAGTGAATATGATGGAAATCTTCGACAGGAGAG 2356  
Qy LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGlu 717  
Db 2357 ATGGATGAGCGATGACGACCTGGATGACAAACGAGCCCATCAAGAGGAGCGCTCC 2416  
Qy AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737  
Db 2417 CCGCGACCGCAGAGCTTCTGCACTCTCTCCAGCATCTCCCCCCAGGAC---AAG 2467  
Qy LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757  
Db 2468 CTCTCATGCCGGGCTTCAGCACGCCAAGGAGCAAGCAACGACTC----- 2512  
Qy LeuGlnAsnGluThrTyrGlyAla-----LeuLeuSerGlySerPro 771  
Db 2513 -----TCTACGGGCGCTTCACCAACCCAGATCTTCGTCTCCACAAGCACAGAC 2560  
Qy ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsn 791  
Db 2561 TCACCCACGTCACCGATCGCA-----GAGGCGCCCGCTCGCTCCCGAAGAAC 2608  
Qy ValGlyLys-----ValGlnThrAlaSer 799  
Db 2609 GCCACGAAGGTCCACCTGGCCCACTTCAACATCTCCCTTAAGCACCCAGACCTTATGT 2668  
Qy SerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819  
Db 2669 GGCAGCTCCACCTGTCCAAGAGCGGTCTCTCTCCCCACCACCCGGA---CACAAGAGA 2725  
Qy SerSerSerAspProProAlaValHisProProLeuProProLeuArg-----ValThr 837  
Db 2726 ACCGTGTGTACCTCTCCAGCCCACTACCTCACGGGCCCCCAAAACAAAGGCGGAGTTCT 2785  
Qy SerThrAsnProLeuThrProThrProProProProValAlaLysThrProSerValMet 857  
Db 2786 TGGGGTAAACGAGTGGGTCCCTCATCGTCC-----AGTAAGACCACCAACAAAGTTC 2836  
Qy GluAlaLeuSerGln----- 862  
Db 2837 GAGGCGCTGTCTCCAGCAGTCGAGCACCGGTTCTGAAAGACTGCATTTGTCCCAAGAGTT 2896  
Qy -----ProSerLysProAla----- 867  
Db 2897 CTTCCTAAACTACCTCAGAAAAGTGGCACTAAGGAAAAACAGAGACCAGCCATCTCTCTCC 2956  
Qy -----ProProGlyIle-----SerGlnIle----- 874  
Db 2957 CTCGACAAAGCAAGTCCCACTCGAGATCTTCAGAAAGTCGTCCAGTTGACAGAGTTA 3016  
Qy -----ArgProProPro-----LeuProProGlnPro----- 883  
Db 3017 CCGCAGAAGCCGACCCCGGGGACCTGCCCCCGAAGCCCAAGGAACTGGCTCCCAAAACC 3076  
Qy -----ProSerArgLeuProGlnLys----- 890  
Db 3077 CCCATTGGAGACTTACCACCTTAAGCCAGCGAGCTGCCCCCGAAGCCACAGCTGGGCGAC 3136  
Qy ----- 890  
Db 3137 CTGCCCCCAAGCCCCAGCTCGCAGACTTGGCCCCCAAGCCCGAGGTGAAGACCTGGCT 3196  
Qy ---LysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909  
Db 3197 CCCAAGCCACAACCTGGGGGAGCTGTGTGCAAAACCCACAGACGGGAGACGCTCGGCCAAG 3256  
Qy Gly-----ProValAspLeuSer----- 915  
Db 3257 GCCCAGGCCACCCCTGGAGCTCACCCCCCAAGTCAACCCCGGGGAGCTGTCTCCCGAACGTC 3316

Qy	916	-----AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnPro	932
Db	3317	CCCAAGCAGGGCGTCTGAGGACACCAACGACCTTCAGCCACCCCTCGCCA---GAGACACCC	3373
Qy	933	AlaProMetProArgLysSerGlnAlaThrIysLeuLysProLysArgValIysAlaLeu	952
Db	3374	GTGCCTCTGCCACGAAATCAACACACGGGGAAGCAAGGTGAGCGCGAGTCAGACCATC	3433
Qy	953	TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIle	972
Db	3434	TACAGCTGCCACGGCGGCAACAGATGACGAGCTGACTTTCATGGAGGCGGAGGTGATCGT	3493
Qy	973	ValAspGlyGluGluAspGlnGluTrpTptIleGlyHisIleAspGlyAspProGlyArg	992
Db	3494	GTACCGGGGAGGAGGACCAGGAGTGTGGATTGGGCACATCGAGGGGCGAGCCGAGAGG	3553
Qy	993	LysGlyAlaPheProValSerPheValHisPheIleAlaAsp	1006
Db	3554	AAGGGCGCTTCCCAAGTGTCCTTTGTCCACATCTCTGCGAC	3595

RESIT.T 7

RESOLUTION /  
ADJ51008  
ID ADJ51008 standard; cDNA; 3974 BP.

PA (STON/) STONE D J.  
PA (TAUP/) TAUPIER R J.  
PA (CASM/) CASMAN S J.  
PA (ROTH/) ROTHENBERG M E.  
PA (MALY/) MALYANKAR U M.  
PA (BOLD/) BOLDOG F L.  
XX  
PI Gorman L, Zerhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R;  
PI Zhong M, Pattrarajan M, Miller CE, Ji W, Pena CE, Burgess CE;  
PI Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME;  
PI Malyankar UM, Boldog FL;  
XX  
WPI: 2004-168942/16.  
DR P-PSDB; ADJ51009.  
XX  
XX  
PT New NOVX polypeptides and polynucleotides, useful in diagnosing, treating  
PT or preventing diseases or conditions, e.g. autoimmune disease,  
PT Alzheimer's disease, diabetes, graft versus host disease, cancer or viral  
PT or bacterial infections.  
XX  
PS Claim 20: SEQ ID NO 73: 342dp: English.

(STON//) STONE D J.  
PA (TAUP//) TAUPIER R J.  
PA (CASM//) CASMAN S J.  
PA (ROTH//) ROTHENBERG M E.  
PA (MALY//) MALYANKAR U M.  
PA (BOLD//) BOLDOG F L.

Gorman L, Zerhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R;  
PI Zhong M, Patturajan M, Miller CB, Ji W, Pena CBA, Burgess CB;  
PI Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME;  
PI Malyankar UM, Boldog FL;  
XX  
WPI; 2004-169942/16.  
DR P-PSDB; ADJ51009.  
DR  
XX

New NOVX polypeptides and polynucleotides, useful in diagnosing, treating  
PT or preventing diseases or conditions, e.g. autoimmune disease,  
FT Alzheimer's disease, diabetes, graft versus host disease, cancer or viral  
or bacterial infections.  
XX  
XX

Claim 20; SEQ ID NO 73; 342pp; English.

The invention relates to an isolated NOVX polypeptide (of 44 disclosed)  
CC comprising its mature form, a sequence having at least 95% sequence  
CC identity to NOVX or a sequence comprising one or more conservative  
CC substitutions in the amino acid sequence of NOVX. Also included are a  
CC composition comprising NOVX and a carrier, a kit comprising, in one or  
CC more containers, the composition, a method of identifying an agent that  
CC binds to NOVX, a method for identifying a potential therapeutic agent for  
CC use in treatment of a pathology related to aberrant expression or  
CC aberrant physiological interactions of NOVX, a method for screening for a  
CC modulator of activity of or of latency or predisposition to a pathology  
CC associated with NOVX, a method for modulating the activity of NOVX, a  
CC method of treating or preventing a pathology associated with NOVX or a  
CC pathological state in a mammal, an isolated nucleic acid molecule  
CC encoding a NOVX protein, a vector comprising the nucleic acid molecule,  
CC a cell comprising the vector, an antibody that immunospecifically binds  
CC to NOVX, a method for determining the presence or amount of NOVX or the  
CC nucleic acid molecule in a sample, a method for determining the presence  
CC of or predisposition to a disease associated with altered levels of  
CC expression of NOVX or the nucleic acid molecule in a first mammalian  
CC subject and a method of producing NOVX (comprising culturing the cell  
CC under conditions that lead to expression of the polypeptide). NOVX is  
CC useful in the manufacture of a medicament for treating a syndrome  
CC associated with a human disease associated with NOVX. The polypeptides  
CC and nucleic acid molecules are useful in diagnosing, treating or  
CC preventing diseases or conditions, e.g. autoimmune disease, Alzheimer's  
CC disease, stroke, allergies, Parkinson's disease, Huntington's disease,  
CC multiple sclerosis, anxiety, pain, diabetes, graft versus host disease,  
CC pancreatitis, obesity, ulcers, anaemia, cancer, viral or bacterial and  
CC parasitic infections (many more diseases and disorders are listed in the  
CC specification). The present sequence encodes a NOVX protein.

XX  
SQ Sequence 3974 BP; 1208 A; 1019 C; 888 G; 859 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.05e-124	Length:	3974
Score:	3006.00	Matches:	626
Percent Similarity:	66.58%	Conservative:	127
Best Local Similarity:	55.35%	Mismatches:	234
Query Match:	57.01%	Indels:	144
DB:	12	Gaps:	22

US-09-914-042-1 (1-1006) x ADJ51008 (1-3974)

Qy 1 MetProAspGlnIleSerValSerGlutPheValAlaGluThrHisGluAspTyrLySAla 20  
Db 321 ATCGCGACAGATCTGTGTCGGAGTTTCATCGCGAGACCACCGAGACTACAACTCG 380  
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValalaAla 40  
Db 381 CCCACCACGTCACGCTTCCACCAACCGCGGTGCACAACCTGCAGGAACACCGGTACGCTGCTG 440



Db 2565 -----TCTATGGAGCCTTACCAACGAGTCTTCGTTTCCACAGCAGAC 2612  
Qy 775 GlnProAlaProSerThrThrSerAlaProProLeuProProAlaAsnValGlyLys 794  
Db 2613 TCGCCACATCACCA---ACCACGAGGCTCCCTCTGCCCCCTAGGAACCGCGGAA 2669  
Qy 795 -----ValGlnThrAlaSerSerAlaAsn 802  
Db 2670 GGTCCAACTGGCCCACTTCAACACTCCTTAAGCACCAGACCTCTAGTGGCAGCTCC 2729  
Qy 803 ThrLeuTpyLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSerSer 822  
Db 2730 ACCCTATCCAGAGAGGCTCTCTCCACACCCCGA---CACAGAGAACCTTATCC 2786  
Qy 823 AspProAlaValHisProProLeuProLeuArg----- 835  
Db 2787 GACCTCCAGCCCACTACCTCATCTGGGCCCCCAACAAAGGCGAGTCTCTTGGGTAAC 2846  
Qy 835 ----- 835  
Db 2847 GATGGGGTCCATCTCTTCAAGTAAGACTACAAAGTTTGGGGACTATCCAGCAG 2906  
Qy 836 -----ValThrSerThrAsnPro---LeuThrProThrProProPro 848  
Db 2907 TCGAGCACCGTCTGCAAGACTGCTCTTGGCCCAAGAGTCTTCTTAAACTACTCAG 2966  
Qy 849 ProValAla-----LysThr-----ProSer 855  
Db 2967 AAGTGGCACTAAGGAAACAGATCATCTCTCCCTAGACAAAGCCACCATCCGCGCGAA 3026  
Qy 856 ValMetGluAlaLeuSerGln-----ProSerLysProAlaPro----- 868  
Db 3027 ATCTTTCAGAAATCATCACAGTGGCAGAGTTGCCCAAAAGCCACCACTGGAGACCTG 3086  
Qy 869 -----ProGlyIleSerGlnIleArgProProPro 878  
Db 3087 CCCCCAAAGCCACAGAACTGGCCCCCAAGCCCCAAATTTGGAGATTTGCCGCTTAAGCCA 3146  
Qy 879 -----LeuProGlnProPro---SerArg 886  
Db 3147 GGAGAACTGCCCCCAACACAGCTGGGGGACCTGCCACCAACCACTCTCAGAC 3206  
Qy 887 LeuProGlnLys----- 890  
Db 3207 TTACCTCCCAACACACAGATGAAGACCTGCCCCCAACACACAGCTGGGAGACCTGCTA 3266  
Qy 891 -----LysProAlaProGlyThrAsp 897  
Db 3267 GCAAAATCCAGACTGGAGATGTCTCACCCAGGCTCAGCAACCTCTGAGGTCACTG 3326  
Qy 898 LysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSer 915  
Db 3327 AGTCACACCATTTGGATCTATCCCAATGTGCAGTCCAGAGCGCCATCCAAAGCAA 3386  
Qy 916 AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuLeuGlnProProAlaProMet 935  
Db 3387 GCATCTGAAGACTCCAACAGCTCAGCGCTACTCTGCCA---GAGACGCCGCTACCACTG 3443  
Qy 936 ProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyranCys 955  
Db 3444 CCGAGAAATCAATACCGGGGAAATTAAGTGGCGGAGTGAAGACCATTTATGACTGC 3503  
Qy 956 ValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGly 975  
Db 3504 CAGGAGAACAGATGACAGCTCATCTTCATTCGAGGAGAGATGATTATCTCAGCGG 3563  
Qy 976 GluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAla 995  
Db 3564 GAAGAGGACAGGAGTGTGGATTGGCCACATCGAAGGACAGCCTGTAAGGAGGGGTC 3623  
Qy 996 PheProValSerPheValHisPheIleAlaAsp 1006  
Db 3624 TTTCAGTGTCTCTTTGTTTCATATCTCTGTCTGAC 3656

## RESULT 8

AAV59104  
ID AAV59104 standard; cDNA; 4382 BP.  
XX  
AC AAV59104;  
XX 17-OCT-2003 (revised)  
DT 20-NOV-1998 (first entry)  
XX Zebrafish differentiation enhancing factor 1 gene.  
XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;  
C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;  
KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;  
nervous system.  
XX Danio rerio.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 351..3806  
FT /\*tag= a  
FT /product= "DEF1 protein"  
XX  
PN WO9836065-A1.  
XX  
PD 20-AUG-1998.  
XX  
PF 13-FEB-1998; 98WO-US002724.  
XX  
PR 14-FEB-1997; 97US-0038191P.  
XX  
PA (DAND ) DANA FARBER CANCER INST INC.  
XX  
PI Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;  
XX  
DR WPI; 1998-467173/40.  
DR P-PSDB; AAW77287.  
XX  
PT New nucleic acid encoding differentiation enhancing factor - used  
PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating  
PT tumours and neurological disease.  
XX  
PS Claim 7; Fig 13; 203pp; English.  
XX  
CC The differentiation enhancing factors (DEF), comprise at least one each  
CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology  
CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they  
CC are mediators of SH3-domain dependent signalling and may be involved in  
CC cellular gene expression, cytoskeletal architecture, protein trafficking,  
CC endocytosis or adhesion, migration, proliferation and differentiation of  
CC cells. Typical applications of DEF and agents that modulate interaction  
CC between the protein and it's ligand, or of nucleic acid expressing them,  
CC are treatment of hyperplastic and neoplastic disease (a wide range of  
CC solid tumours and leukaemias), including metastases; for in vitro  
CC induction of differentiation of neural crest cells to neurons, glial  
CC cells etc.; for increasing neuron survival, and inducing cell repair, in  
CC the nervous system (e.g. treatment of traumatic injury, stroke, in  
CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral  
CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 4382 BP; 1222 A; 1132 C; 1098 G; 930 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 1,46e-121 Length: 4382  
Score: 2954.50 Matches: 622  
Percent Similarity: 65.04% Conservative: 126  
Best Local Similarity: 54.09% Mismatches: 239  
Query Match: 56.03% Indels: 163  
DB: 2 Gaps: 24  
US-09-914-042-1 (1-1006) x AAV59104 (1-4382)

```
Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
Db 411 ATGCCGGATCAGATCTCCGTCGTCGAGTTCTCTCGAGAGCAGCGAGGATTAACAATTC 470
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 471 CCCACGACCTCGAGCTTCCACCACCGCTCGAGAGCTGCGGAAACACCGTCAATGTTCTG 530
Qy 41 GluGluAlaLeuAppValAspArgMetValLeuTyrIysMetLysLysSerValIysAla 60
Db 531 GAAGAGGCTTTGGATCAGACCGCACTGCTTTACAGAAGGTCAAGAAATCTGTCAAAGCA 590
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 591 ATCTACAACCTCGGGTCAAGACATGTGCAGAAATGAAGAAATATTGGACGACCTGGAG 650
Qy 81 LysPheGlyGlyAsnCysValCysArgAspPheProaspLeuGlyIserAlaPheLeuIys 100
Db 651 AAGTTTGCAGCAACTTTCATCAGCCGAGATAACTCTGATCTGGGAACAGCCTTCATCAAG 710
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 711 TTTTCTGGACTTATCAAAGAGCTGCTGCTCTCTCAAAGAACTGCTCCAGAGCCTCAGC 770
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
Db 771 CACAACGTCATCTTCCCTCGACTCTCTGCTCAAAGGAGATCTTAAGGGAGTGAAGGGG 830
Qy 141 AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrIysIle 160
Db 831 GACCTTAAAGCCCTTTCGACAGCCCTGGAAGACTATGAACCAAGTTTCAAAAGATC 890
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 891 GAGAAGGAGAAGAGAAACATGCCAAGCAGCAGCGCATGATCCGCACAGAAATCACCGGC 950
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 951 GCAGAGATTGAGAAAGAGATGAGAAAGAGCGGAGGATCTTTACGCTGCAGATGTGTAG 1010
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 1011 TACCTGATCAAGTCAATGAGATTAAAGCAAGAGGGAGTGGATCTCTCCAGATCTC 1070
Qy 221 IleIysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1071 ATCAAGTATTATCATGCACAGTGCATTTCTTCCAGGATGGCTTGAAAACTGCTGACAAG 1130
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1131 TTGAAGCAGTATATTGAAAAATTTAGCAGCTGATCTTTATATATATAAAACAGACTCAGGAT 1190
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1191 GAGGAGAAAAACAGCTCACAGCTCTCAGAGACCTCATCAATCTTCTTACAGCTGGAC 1250
Qy 281 GlnLysGluAppSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1251 CAGAAGGAGGATCTCAGAGTAAGCAGAGC---GGGTACAGCATGCACCAGCTGCAGGCG 1307
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1308 AATAGGAGTTTGGCAGTGAGAAGAGGGCTATCTCTTCAAAGAGAGTGTGGGATCCGT 1367
Qy 321 LysValTyrGlnLysArgLysCysSerValIysAsnGlyPheLeuThrIleSerHisGly 340
Db 1368 AAGTGTGCAGAGAGGAGAGTGTCTCAGTGAANAATGSCATCCTCACCATCTCTCATGCC 1427
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValIysThrAspPro 360
Db 1428 ACATCCAAACAGCAGCCGTCGAGACTGAATCTGCTGACTGCGCAGGTAAACCCAGTGA 1487
```

---

```
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1488 GAGATAGAGAGTCTCTTTGACCTCATCTCTCATTAATCGACATATCATTTCCAGGCAGAG 1547
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1548 GACGAACAGGAGTTTGTGATATGATCTCGGTGCTGACTAATAGTAGAAGGAGGCTCTG 1607
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1608 AACATGTCATTTCTGGGGGAGAGAGTCTGAGATGACAGTTTGG---GAGACTTTGACC 1664
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1665 AAAGCCATCATCGAGGAGTGTCTCGCATTCCTGGAAACGNAAGTCTGCTGTGACTGTGGG 1724
Qy 441 AlaProaspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1725 GTTCCAGAGCCCAAAATGGTTATCCACTAACTCGGCATCTCTGAGGTGCATCGAGTGTTC 1784
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1785 GGAATCCACAGGGAATGGGAGTCCATATTTCCGCGCATCCAATCCATGAGCTTGACAAA 1844
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1845 CTTGGAACCTCTGAACCTCTTGTGCTTAGNACGTGGCAACAGTAGTTTCAACGAATA 1904
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1905 TTAGAAGGAATCTGCCAGT---CCTTCAACAAGCCAGGCCCATCAAGTGACATGACC 1961
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1962 GAGAGAGGAGTACATCAATCGGAAGTACGTGAGCAGCAGGTTTCGTCCGCGACCGCC 2021
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 2022 ACTACAGCCACAGCCACAGCGGCGACTTGTACAGGCGGTGAGNACCGGAGACTTGATG 2081
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2082 GCTCTCTTACGCTCTATGCAGATGGAGTGAGCTAATGGATCTCTTCCACAGAAGCA--- 2138
Qy 581 GlyHisGluProaspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2139 GGACAGACCGCGGAGAGACAGCTCTGACTTTCTGTCGACATCATCAGACCATCAGACACTTC 2198
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2199 CTGCACTGTGTGGACTTCTTGTCCAAAACAGTGGGACTCTAGACAGACACCGGAGAGT 2258
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2259 GGAACCGTCTCTCTCCATCTACTGTCACATATGAGAGCCAGAGTGTCTCAAACTGCTG 2318
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2319 CTCAGGGGAAAACCGTCTATTGACTGTGTTAATCAAAACCGGGAGACAGCATTTGGATATC 2378
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2379 GCCAGACATGAGAAATGTACAGTGTGAAGAGTACTGTGTGGAGCGACAGCCGGGAGG 2438
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2439 TTTAATCTCATGTGCATGTGAGTATGAGTGAATCTCGCGCTGGAGGAGATGATGAG 2498
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2499 AGTACCATGATCACCTGGATGACAGCCCTAGTCCAGTGAAGAGGAGCGTTCTTCCTCGTCT 2558
Qy 721 IleSerPheTyrGlnLeuLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
```









```
Db 1776 ACATCTCCATTCGTTGACTTCCTGTACAAAACCTGGGAACCTGGATTAAGCAGCG 1835
Qy |||||
Db 619 GlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLys 638
Qy |||||
Db 1836 GCCCTGGGAACACAGATTCACACTACTGTAGTATGTACAGTAAACCTGAGTGTITGAAG 1895
Qy |||||
Db 639 LeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeu 658
Qy |||||
Db 1896 CTTTTCCTCAGAGACAGCCCACTGTGATATAGTTAAACAGGCTGGAGAACTCCCTCA 1955
Qy |||||
Db 659 AspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGluAlaLeuSer 678
Qy |||||
Db 1956 GACATAGCAAGAGACTTAAGCTACCCAGTGTGAAGATCTGCTTTCCAGGCTAAATCT 2015
Qy |||||
Db 679 GlyArgPheAsnSerHisValHisValGluTyrGluTyrArgGluLeuHisAspLeu 698
Qy |||||
Db 2016 GGAAGTTCAATCCACAGCTCCACGTAGATATAGTGGAAATCTTCGACGAGGAGATA 2075
Qy |||||
Db 699 AspGluSerAspAspMetAspGluLysLeuGluGlnProSerGluAsnArgGluAsp 718
Qy |||||
Db 2076 GATGAGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2135
Qy |||||
Db 719 ArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeu 738
Qy |||||
Db 2136 AGACCTCAGAGCTTCTGCACCTCTCCAGCATCTCCCCCGAGGAC-----AAGCTG 2186
Qy |||||
Db 739 AlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeu 758
Qy |||||
Db 2187 GCATCGCCAGGATTCAGCACTCCAGGAGCAACACAGCGCTC----- 2228
Qy |||||
Db 759 GlnAsnGluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGln 775
Qy |||||
Db 2229 -----TCTATGGAGCTTCCACCAACAGCATCTTCGTTTCCACCAAGCAGACTCG 2279
Qy |||||
Db 776 ProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys--- 794
Qy |||||
Db 2280 CCCATCACCA-----ACCACGAGGCTCCCTCTGCTCTCTAGGAACCGCGGAAAGGT 2336
Qy |||||
Db 795 -----ValGlnThrAlaSerSerAlaAsnThr 803
Qy |||||
Db 2337 CCAACTGGCCACCTTCAACTCCCTTAAGCACCAGACTCTAGTGGCAGCTCCACC 2396
Qy |||||
Db 804 LeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSerSerSerAsp 823
Qy |||||
Db 2397 CTATCCAAAGAGAGGCTCTCTCCCAACCAACCGCA-----CACAGAGAACCTTATCCGAC 2453
Qy |||||
Db 824 ProProAlaValHisProProLeuProLeuArg----- 835
Qy |||||
Db 2454 CTTCCAGCCCACTACCTCATCTGGGCCCCCAACAAAGCGCAGTTCTCTTGGGTAACGAT 2513
Qy |||||
Db 835 ----- 835
Qy |||||
Db 2514 GGGGTCCATCTCTTCAAGTAAGACTACAAACAGTTTGGAGGACTATCCAGCAGTCG 2573
Qy |||||
Db 836 -----ValThrSerThrAsnPro-----LeuThrProThrProProPro 849
Qy |||||
Db 2574 AGCACCAGTCTCGAAAGACTGCTTGGCCCAAGAGTCTCTCTAACTACTACCTCAGAA 2633
Qy |||||
Db 850 ValAla-----LysThr-----ProSerVal 856
Qy |||||
Db 2634 GTGGCACTAAGGAAACACAGATCATCTCTCTAGACAAAGCCACCCTCCCGCGGAAATC 2693
Qy |||||
Db 857 MetGluAlaLeuSerGln-----ProSerLysProAlaPro----- 868
Qy |||||
Db 2694 TTTCAGAAATCATCAGTTGGCAGAGTTGGCCACAAAGCCACCACCTCCGAGACTGCCC 2753
Qy |||||
Db 869 -----ProGlyLysSerGlnIleArgProProProPro--- 878
Qy |||||
Db 2754 CCAAGCCACAGAACTGGCCCCCAGGCCCAATTTGAGATTTGCCCTTAAGCCAGGA 2813
Qy |||||
Db 879 -----LeuProProGlnProPro---SerArgLeu 887
Qy |||||
```

```
Db 2814 GAACCTGCCCCCAACACACAGCTGGGGGACCTGCCCAACCCCAACTCTCAGACTTA 2873
Qy |||||
Db 888 ProGlnLys----- 890
Qy |||||
Db 2874 CCTCCAAACACACAGATGAAGACCTGCCCCCCCAACACACAGCTGGGAGACCTGTAGCA 2933
Qy |||||
Db 891 -----LysProAlaProGlyThrAspLys 898
Qy |||||
Db 2934 AAATCCCAAGCTGGAGATGTCTACCAAGGCTCAGCAACCTCTTGAGGTCACTGAAG 2993
Qy |||||
Db 899 SerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAla 916
Qy |||||
Db 2994 TCACACCCATCTGGATCTATCCCAATGTGCTCCAGAGCGCCATCCAAAGCAAGCA 3053
Qy |||||
Db 917 ThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetPro 936
Qy |||||
Db 3054 TCTGAAGACTCCAAGACCTCAGCTACTCTGCCA---GAGACGCCGTACCTGCC 3110
Qy |||||
Db 937 ArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysVal 956
Qy |||||
Db 3111 AGAAAAATCAATACGGGGAAAAATAAAGTGGCGAGTGAAGACCATTTATGACTGCCAG 3170
Qy |||||
Db 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGlu 976
Qy |||||
Db 3171 GCAGACAACGATGACGAGCTCACATTATCATCGAGGAGAAAGTATTATCTACAGGGGAA 3230
Qy |||||
Db 977 GluAspGlnGluTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Qy |||||
Db 3231 GAGNACCAAGAGTGGTGGATTGGCCCATCATCAAGGACAGCTTGAAGGAGGGGGTCTTT 3290
Qy |||||
Db 3291 CCAGTGTCTTTGTTCATATCTGTCTGAC 3320
Qy |||||

RESULT 10
ADS10088
ID ADS10088 standard; DNA; 4997 BP.
XX
AC ADS10088;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human therapeutic DNA - SEQ ID 325.
XX
KW antiinflammatory; neuroprotective; antianaemic; cytosstatic; vulnerary;
XX inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
XX aplastic anaemia; cancer; wound healing; gene therapy; db; gene.
XX
OS Homo sapiens.
XX
PN WO2004080148-A2.
XX
PD 23-SEP-2004.
XX
PF 30-SEP-2003; 2003WO-US030720.
XX
PR 02-OCT-2002; 2002US-0416186P.
XX
PA (NUVE-) NUVELO INC.
XX
PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y; Zhou P;
PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX
DR WPI; 2004-668857/65.
XX
DR P-PSDB; ADS10772.
XX
PT New polynucleotide, useful in preparing a composition for diagnosing or
PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
XX aplastic anemia or cancer for promoting wound healing.
XX
PS Claim 1; SEQ ID NO 325; 718pp; English.
XX
CC The invention relates to a novel isolated polynucleotide and the encoded
```



Qy 711 oSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731  
Db : : : : :  
1799 TATCAAGAAAGAGCGCTCACCAGACCTTCAGAGCTTTCGCCACTCTCCACAGCTCTCCCC 1858  
Qy 731 uGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArg 751  
Db : : : : :  
1859 CCAGGAC-----AAGCTGGCACTCCAGGATTACAGACTCCAGGGACAAACAGCG 1909  
Qy 751 gAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly---- 769  
Db : : : : :  
1910 GCTC-----TCCTATGGAGCTTCACCAACACAGATCTT 1942  
Qy 770 ----SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPr 788  
Db : : : : :  
1943 CGTTTCCACAAGCAGACTCGCCACATACCA---ACCACGGAGGCTCCCTCTCTGCC 1999  
Qy 788 oProArgAsnValGlyLys-----ValG1 796  
Db : : : : :  
2000 TCCTAGGAACCGCGGAAGGTCCCACTGGCCCACTTCAACACTCCCTCTAAGCACCCA 2059  
Qy 796 nThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyLys 816  
Db : : : : :  
2060 GACTCTAGTGGCAGCTCCACCTATCCAAAGAGAGGCTCTCTCCCCACCAACCGGA-- 2117  
Qy 816 rArgGlnArgSerSerAspProProAlaValHisProProLeuProProLeuArg-- 835  
Db : : : : :  
2118 -CAACAAGAGAACCTATATCCGACCTCCAGCCCACTACTCTATGGGGCCCCCAACAAGG 2176  
Qy 835 ----- 835  
Db 2177 CGCAGTTCCTTGGGGTAAGATGGGGTCCATCTCTTCAAGTAAGACTACAAACAGTT 2236  
Qy 836 -----ValThrSerThrAsnPro---Le 842  
Db : : : : :  
2237 TGAGGGACTATCCAGCAGTCGAGCACCAGTTCTGCAAGACTGCCCTTGGCCCAAGAGT 2296  
Qy 842 uThrProThrProProProValAla-----LysThr----- 853  
Db : : : : :  
2297 TCTTCTAAACTACTCTCAGAAAGTGGCACTAAGGAAACAGATCATCTCTCCCTAGACAA 2356  
Qy 854 -----ProSerValMetGluAlaLeuSerGln-----ProSerLy 865  
Db : : : : :  
2357 AGCCACCATCCCGCCCGAATCTTTCAGAAATCATCATGTTGGCAGAGTTGCCACAAA 2416  
Qy 865 sProAlaPro-----ProGlyIleLe 872  
Db : : : : :  
2417 GCCACCACTGGAGACTGCCCCCAAGCCACAGACTGCGCCCAAGCCCAAAATGG 2476  
Qy 872 rGlnIleArgProProPro-----LeuProPr 881  
Db : : : : :  
2477 AGATTGCGGCTAAGCCAGGAGAACTGCCCCCAACACAGCTGGGGGACCTGCCACC 2536  
Qy 881 oGlnProPro---SerArgLeuProGlnLys----- 890  
Db : : : : :  
2537 CAACCCCAACTCTCAGACTTACTTCCCAACCAACAGATGAAGACTGCCCCCAACACC 2596  
Qy 891 -----Ly 891  
Db 2597 ACAGCTGGAGACCTGTAGCAAAATCCAGACTGGAGATGCTCTCACCACAGGCTCAGCA 2656  
Qy 891 sProAlaProGlyThrAspLysSerThrPro-----LeuThrAsnLysGlyGlnProAr 909  
Db : : : : :  
2657 ACCCTCTGAGTGCACACTGAAGTCAACCCATTCATCCCAAAATGTGAGTCCAG 2716  
Qy 909 gGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLe 929  
Db : : : : :  
2717 AGCGCCATCAAAAGCAAGCATCTGAAGACTCCCAACGACTCCAGCTACTCTGCCA-- 2774  
Qy 929 uGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVa 949  
Db : : : : :  
2775 -GAGACGCCCTACACTGCTCCAGAAAATCAATACGGGGGAAAATAAAGTGGCGAGT 2833  
Qy 949 lLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAs 969

Db 2834 GAAGACCATTTATGACTGCCAGGCAGACACGATGACGAGCTCACATTCATCGAGGGAGA 2893  
Qy 969 pValIleIleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAs 989  
Db : : : : :  
2894 AGTATATATCTCAGGGGAGAGGACGAGAGTGGTGGATTGGCCACATCGAAGGACA 2953  
Qy 989 pProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaApp 1006  
Db : : : : :  
2954 GCCTGAAGGAAGGGGTCTTTCAGTGTCTCTTGTTCATATCTCTGTCTGAC 3005  
RESULT 11  
AAI59224  
ID AAI59224 standard; cDNA; 5033 BP.  
XX  
AC AAI59224;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 1427.  
XX  
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR P-PSDB; AAM40068.  
XX  
FT Novel nucleic acids and polypeptides, useful for treating disorders such  
as central nervous system injuries.  
XX  
PT Claim 1; SEQ ID NO 1427; 10078pp; English.  
XX  
PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nontropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and





CC useful for identifying an agent that modulates an activity of a POSH  
 CC polypeptide or POSH-AP, for identifying an antiviral agent, an anti-  
 CC apoptotic agent, an anti-cancer agent, an agent that inhibits trafficking  
 CC of a protein through the secretory pathway, an agent that inhibits the  
 CC progression of a neurological disorder, an agent that modulates a POSH  
 CC function, an agent that modulates a HERPUD1 function. The methods can be  
 CC used for treating a viral infection, for inhibiting an activity of a POSH  
 CC -AP in a cell, for treating a POSH-associated disease in a subject. The  
 CC POSH-associated disease is viral infection, POSH-associated cancer or  
 CC POSH-associated neurological disorder. The methods are useful for  
 CC treating or preventing POSH-associated neurological disorder in a subject  
 CC e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease,  
 CC schizophrenia, Niemann-Pick's disease. This sequence corresponds to a  
 CC nucleic acid of the invention.

XX SQ Sequence 4870 BP; 1509 A; 1159 C; 997 G; 1205 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 6,36e-100 Length: 4870  
 Score: 2465.50 Matches: 520  
 Percent Similarity: 64.62% Conservative: 110  
 Best Local Similarity: 53.33% Mismatches: 204  
 Query Match: 46.76% Indels: 141  
 DB: 13 Gaps: 21

US-09-914-042-1 (1-1006) x ADS34272 (1-4870)

Qy 154 GluThrIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHisGlyMet 173  
 Db 1 GAGACAAAGTTTACAAAATTCAGAAAGAGAAAGAGAGACGCGCAAAACACATGGGATG 60  
 Qy 174 IleArgThrGluIleSerGlyValaGluIleAlaGluMetGluLysGluArgArgPhe 193  
 Db 61 ATCCGCACAGAGATAACAGGAGCTGAGATTGCGAAGAAATGGAAGAAAGGCGCTC 120  
 Qy 194 PheGlnLeuGlnMetCysGluThrLeuLeuLysValAsnGluIleLysIleLysGly 213  
 Db 121 TTTGAGTCCCAATGTGTGAATATCTCATTAAAGTTTAAAGTTTAAATCAACACCAAGGGT 180  
 Qy 214 ValAspLeuLeuGlnAsnLeuLysTyrPheHisAlaGlnCysAsnPhePheGlnAsp 233  
 Db 181 GTGGATCTGCTGCAGAACTTATATAAGATTATACCATGCACAGTGCATTTCTTTCAAGAT 240  
 Qy 234 GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis 253  
 Db 241 GCGTTGAAACAGCTGATAAGTTGAACAGTACATTGAAACACTGGCTGCTGATTATAT 300  
 Qy 254 ThrIleLysGlnAlaGlnAspGluArgGlnLeuIleGlnLeuArgAspIleLeu 273  
 Db 301 AATATAAACAGACCCAGGATGAAGAAAGAAACAGCTAATGCACTCCGAGACTTAATA 360  
 Qy 274 LysSerAlaLeuGlnValGluLysGluAspSerGlnIleArgGlnSerThrAlaTyr 293  
 Db 361 AAATCCTCTCTCAACTGGATCAGAAAGAAAGATTCTCAGAGCGCGGGA---GGATAC 417  
 Qy 294 SerLeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyr 313  
 Db 418 AGCATGCATACGCTCCAGGGCAATGAAGAAATATGGCAGTGAAGAAAGGGGTACCTGCTA 477  
 Qy 314 LysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGly 333  
 Db 478 AAGAAAGTACGGGATCCGAAAGTATGGCAGAGAGGAGGAGTGTTCAGTCAAGAAATGG 537  
 Qy 334 PheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThr 353  
 Db 538 ATTTCTGACCATCTCATATGCCACATCTAACAGGCAACCCAGGCAAGTGTAAACCTTCTACC 597  
 Qy 354 CysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHisAspArg 373  
 Db 598 TGCCAAAGTAAACCTTAATGCCGAAGACAAAAAATCTTTTGACCTGATATCATATAAGA 657  
 Qy 374 ThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGln 393  
 Db 1729 -----AAGCTGGCACTGCCAGGATTCAGCACTCCAGGGGCAAAACAGGGGCTC--- 1776

Db 658 ACATATCACTTTTACGGCAGAGATGAGCAGGATTATTGTAGCATGGATATCATGATTGACA 717  
 Qy 394 AsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsn 413  
 Db 718 AATAGCAAGAGAGAGCCCTAAACCATGGCTTCGTGGAGAGCAGAGTGGCGGAGAGAAC 777  
 Qy 414 AsnIleValGlnLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsn 433  
 Db 778 AGCTG---GAAGACCTGACAAAGCCATTATTAGGATGTCACGGCTCCAGGGAT 834  
 Qy 434 AspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIle 453  
 Db 835 GACATTTGCTGGATTGTGGCTCATCAGAACCCACCTGGCTTTCAACCAACTGGGTATT 894  
 Qy 454 LeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMet 473  
 Db 895 TTGACCTGTATAGAATGTTCTGGCATCATAGGAAATGGGGTTTCATATCTCTCGCAAT 954  
 Qy 474 GlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGly 493  
 Db 955 CAGTCTTTGGACTAGACAAATTTAGGAACCTTCTGAACTCTGCTGGCCAGAAATGTAGA 1014  
 Qy 494 AsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysPro 513  
 Db 1015 AACAAATAGTTTTAATGATATTATGGAAGCAAAATTTACCAGC---CCCTCACCAAAACC 1071  
 Qy 514 AsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533  
 Db 1072 ACCCTTTCAAGTATGATGACTGTACGAAAGAAATATATCACTGCAAAAGTATGTAGATCAT 1131  
 Qy 534 ArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLysGluAla 553  
 Db 1132 AGTTTTCAAGGAAGACCTGTTCAACTTCATCAGCTAAACTAAATGAATGCTTTGAGGCC 1191  
 Qy 554 ValLysThrArgAspIlePheGlyLeuGlnAlaTyrAlaAspGlyValAspLeuThr 573  
 Db 1192 ATCAAAATCCAGGGATTACTTTGCACATAATCAAGTCTATGCAAGAGGGGTAGAGCTAATG 1251  
 Qy 574 GluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaVal 593  
 Db 1252 GAGCCACTG---CTGGAACCTGGCAGAGCTTGGGAGAGACAGCCCTTCACCTTCCCGTC 1308  
 Qy 594 ArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsn 613  
 Db 1309 CGAACTGCAGATCAGACATCTCTCCATTTGTTGCTTCTTGTACAAACCTGTGGGAAC 1368  
 Qy 614 LeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsn 633  
 Db 1369 CTGGATAAGCAGACGGCCCTGGGAACACACAGTTCTACACTACTGTAGTATGTACAGTAAA 1428  
 Qy 634 AlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSer 653  
 Db 1429 CCTGAGTGTTTGAAGCTTTTGTCTCAGGAGCAAGCCCACTGTGGATATAGTTAAACAGGCT 1488  
 Qy 654 GlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeu 673  
 Db 1489 GGAGAACTGCCCTAGACATAGCAAGAGACTAAAGCTACCCAGTGTGAAGATCTGCTT 1548  
 Qy 674 ThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeu 693  
 Db 1549 TCCAGGGCTAAATCTGGAAAGTTCAATCCACACGCTCCACGCTAGATAATATGAGTGGAACTT 1608  
 Qy 694 LeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu 713  
 Db 1609 CGACAGGAGGAGATAGATGAGAGCGATGATCTGGAATGACAAACCAAGCCCTTATCAAG 1668  
 Qy 714 AsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSer 733  
 Db 1669 AAAGAGCGCTCACCCAGACCTCAGAGCTTCTGCCACTCTCCAGCATCTCCCCCAGGAC 1728  
 Qy 734 AsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPhe 753  
 Db 1729 -----AAGCTGGCACTGCCAGGATTCAGCACTCCAGGGGCAAAACAGGGGCTC--- 1776





CC progression of a neurological disorder, an agent that modulates a POSH  
 CC function, an agent that modulates a HERPUD1 function. The methods can be  
 CC used for treating a viral infection, for inhibiting an activity of a POSH  
 CC -AP in a cell, for treating a POSH-associated disease in a subject. The  
 CC POSH-associated disease is viral infection, POSH-associated cancer or  
 CC POSH-associated neurological disorder. The methods are useful for  
 CC treating or preventing POSH-associated neurological disorder in a subject  
 CC e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease,  
 CC schizophrenia, Niemann-Pick's disease. This sequence corresponds to a  
 CC nucleic acid of the invention.

XX  
 SQ Sequence 5475 BP; 1680 A; 1264 C; 1086 G; 1445 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2.02e-97 Length: 5475  
 Score: 2410.00 Matches: 511  
 Percent Similarity: 64.15% Conservative: 110  
 Best Local Similarity: 52.79% Mismatches: 203  
 Query Match: 45.70% Indels: 144  
 DB: 13 Gaps: 22

US-09-914-042-1 (1-1006) x ADS34273 (1-5475)

Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIle 183  
 Db 2 AAAAGAGAGCAGCGAAACACATGGGATGATCCGACAGAGATTAACAGGAGCTGAGATT 61  
 Qy 184 AlaGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGluTyrLeuLeu 203  
 Db 62 GCGGAAGAAATGGAGAGGAAAGGCGCTTTTCAGCTCCAAATGTGTGAATATCTCAT 121  
 Qy 204 LysValAsnGluIleLysIleLysGlyValAspLeuLeuGlnAsnLeuIleLysTyr 223  
 Db 122 AAAGTTAATGAATCAAGACCAAAAGGTTGGATCTGCTGCGAATCTTTATAAGTAT 181  
 Qy 224 PheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243  
 Db 182 TACCATGCACAGTGCATTTCTTCAAGATGGCTTGAACACAGCTGATGAAGTGAACAG 241  
 Qy 244 SerLeuGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluArg 263  
 Db 242 TACATTGAAATCGCTGCTGATTTATATATATAATAAACAACAGCCAGGATGAAGAAAG 301  
 Qy 264 ArgGlnLeuIleGlnLeuArgAspIleLysSerAlaLeuGlnValGluGlnLysGlu 283  
 Db 302 AAACAGCTAACTGCATCCGAGACTTAATAAAATCTCTTCAACTGGATCGATCAGAAAGAA 361  
 Qy 284 -----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300  
 Db 362 TCTAGGAGAGATTCTCAGAGCGGCAAGGA---GGATACAGATGCATCAGCTCCAGGGC 418  
 Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320  
 Db 419 AATAAGGAATATGGCAGTGAAGAAAGGGGTACCTGTCTAAAGAAAGTGAACGGGATCCGG 478  
 Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340  
 Db 479 AAAGTATGCGAGAGAGAGAGTGTTCAGTCAAGAAATGGGATTTCTAACCATCTCCATGCC 538  
 Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360  
 Db 539 ACATCTACAGCAGCAACCGCCAGTTGAACCTTCTCACCTGCCAAGTAAACCTTANTGCC 598  
 Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380  
 Db 599 GAAGACAAAAAATCTTTTACCTGATATACATAATAGAACATATCACTTTTCAGGCAGAA 658  
 Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400  
 Db 659 GATGACGAGGATATTGTAGCTGGATATACAGTATTGACAAATAGCAAAAGAGAGGCCCTA 718  
 Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnLeuLeuThr 420

Db 719 ACCATGGCTTCCGTGAGAGACAGAGTGGGGAGAGAAACAGCCTG---GAAGACCTGACA 775  
 Qy 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440  
 Db 776 AAACCCATTATTGAGGATGTCACGCGCTCCACGGGAATGACATTTGCTGGATTGTGGC 835  
 Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysValIleGluCysSer 460  
 Db 836 TCATCAGAACCCCTGCTTTCAACCACTTGGGTATTTTACCTGTATAGAAATGTTCT 895  
 Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480  
 Db 896 GGCATCCATAGGGAATGGGGTTTCATATTCTCGCATTCAGTCTTTTGAACATAGACAA 955  
 Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500  
 Db 956 TTAGAACTTCTGAACTCTTGTGCGCAAGAATGTAGAAACAATAGTTTTTAATGATATT 1015  
 Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520  
 Db 1016 ATGAAGCAAAATTTACCAGC---CCCTCACCACAAACCCACCTTCAAGTGATATGACT 1072  
 Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHis 540  
 Db 1073 GTACGAAAGAAATATATCACTGCAAGATATGTAGATCATAGTTTTCAGGAAGACCTGT 1132  
 Qy 541 AlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560  
 Db 1133 TCAACTTCTCAGCTAAACTAAATGAATGTGTGAGGCCATCAATCCAGGATTTACTT 1192  
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580  
 Db 1193 GCATTAATCAAGTCTATGCAAGGGGTAGAGCTAATGGAAACCACTG---CTGAACCT 1249  
 Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600  
 Db 1250 GGGCAGGAGCTTGGGGAGACAGCCCTTCACTTCCGCTCGCAACTGCAGATCAGACATCT 1309  
 Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620  
 Db 1310 CTCCATTGTGGTGTACTTCTTGTGTAACAACTGTGGGAACCTGGGATAAGACACAGCCGCTG 1369  
 Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640  
 Db 1370 GGAACACAGCTTCTACACTACTGTAGTATGTACAGTAAACCTGAGTGTGTGAAGCTTTG 1429  
 Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660  
 Db 1430 CTCAGGAGCAAGCCCACTGTGGATATAGTTAAACAGGCTGGAGAACTGCCCTAGACATA 1489  
 Qy 661 AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680  
 Db 1490 GCAAGAGACCTAAAGAGTACCCAGTGTGAAGATCTGCTTTCCAGGCTTAAATCTGGAAG 1549  
 Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700  
 Db 1550 TTCAATCCACAGCTCCACGTAGATATGAGTGAATCTTCCACAGGAGGAGATAGATGAG 1609  
 Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720  
 Db 1610 AGCGATGATGATCTGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCACAGACT 1669  
 Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740  
 Db 1670 CAGAGCTTCTGCCACTCTCCAGCATCTCCCCCAAGGAC-----AAGCTGGCACTG 1720  
 Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760  
 Db 1721 CCAGATTACGACTCCNAGGGACAAACAGCGGCTC----- 1756  
 Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla 777  
 Db 1757 ---TCCTATGGAGCTTCCACCAACAGATCTTCGTTTCCCAACAGACAGACTCGGCCACA 1813





## Alignment Scores:

Pred. No.: 2,01e-83 Length: 3812  
 Score: 2089.50 Matches: 455  
 Percent Similarity: 59.20% Conservative: 153  
 Best Local Similarity: 44.30% Mismatches: 275  
 Query Match: 39.63% Indels: 145  
 DB: 10 Gaps: 19

US-09-914-042-1 (1-1006) x ADB62827 (1-3812)

Qy 1 MetProAspGlnIleSerValserGluPheValAlaGluThrHisGluAspTyrHisAla 20  
 Db ATGCGGAGGAGTTCAGCGTGCAGGTTCTTGGCGGTCCCGCGGAGGAGCTTCAGCTCC 106  
 Qy 21 ProThr---AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 39  
 Db CCGGTGGGGCGCGCGCTTCGCCGCCAAGATGCCCGGTACCGAGGGCGCGCGTGGCG 166  
 Qy 40 IleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLys 59  
 Db CGGGAGGAGATCTTGGAGGAGACCAAGCCATCTTCGACAGAAATAAAGAGCTGTGCGG 226  
 Qy 60 AlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeu 79  
 Db GCAATCCATAGTCCGGCGCTTGGCCATGTGGAGATGAAGAGCAGTACCGAGAGGCCGTG 286  
 Qy 80 GluLysPheGlyAsnCysValCysArgAspAspProAspLeuLysSerAlaPheLeu 99  
 Db GAATCTTTAGGCAACAGACACCTGTCCAGAACACAGCCATGAGCTTCCACAGGCTTCCTA 346  
 Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMet 119  
 Db AACTTGGCGGCTTCCCGCGAGGTGCTGCGCTCTTCAGAACTGATTCAGAACTTG 406  
 Qy 120 AsnAsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLys 139  
 Db AACAACTTGTCTTTTCCCTGACAGTCTGATGAAGGGGCGAGCTGAGGGAGCGTCA 466  
 Qy 140 GlyAspLeuLysLysPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLys 159  
 Db CAGGATTCCAAAAAACAGCTGGAGAGGAGTGAAGAGACTATGAAGCCAAATGGCCAA 526  
 Qy 160 IleGluLysGluLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179  
 Db CTGGAGAGGAG---CGGATCGCGCCAGGTGACAGAGGATC----- 568  
 Qy 180 GlyAlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCys 199  
 Db CTGGGAGGTGGCCAGACATGCAGAGAGAGCGCGCATCTTCCAGCTGCACATGTGT 628  
 Qy 200 GluTyrLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219  
 Db GAGTATCTGCTCAAGCCGGGAGAGCCAGATGAAGCAAGGTCCTGACTTCCCTTCAGAGC 688  
 Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu 239  
 Db CTCATCAAGTCTTCCAGCCCGAGCACAACCTTTTCCAAAGTGGCTGGAAGGCTGCCCAG 748  
 Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259  
 Db AGCCTGTTCCTTCATCGAAGAGTGGCGGCTCAGTACATGCATCCATCAGGCCCGAG 808  
 Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279  
 Db GAGGACGAGCTACAGAAGCTGACCCAGCTCCGGGACTCCCTCCGAGGGACACTGCAGCTT 868  
 Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297  
 Db GAGACGACGAGAGAAACCTGAGCCGAGAGAACTCAGGATGTGGCTATAGCATCCACCAG 928  
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317  
 Db CACCAAGGCAACAAGCAGCTTTGGGAGCGGAGAAAGTGGGCTTTCTATACAAGAAAGTGC 988

Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337  
 Db GGAATTCGAGAGTCTGCGAGAAAGGAAGTGTGGAGTCAAGTATGGCTGCTGACCATC 1048  
 Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357  
 Db TCACACAGCAGATAAACCCGCCCGGTGAAGCTGACCCCTGCTGACGTGCCAAGTAGG 1108  
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377  
 Db CCAAAACCTGAGGAGAAAAAGTCTGACCTGCTGACCCACCAACCGGACGTACCACTTT 1168  
 Qy 378 GlnAlaGluAspGluGlnLysCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397  
 Db CAGCAGAGGACGACGAGTGTGAGCGTGGGTGTGAGTGTTCAGTGTTCAGAAACAGAGAG 1228  
 Qy 398 GluAlaLeuAsnAlaPheLysGlyAsp-----AspAsn 409  
 Db GAAGCCCTGAGCAGCGCTTCTCGGGGAGCCAGCGCTGGCCCGGGGTCTCTGGGGGTCC 1288  
 Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleIleSerGluValGlnArg 429  
 Db GCCGCGCATGATGGGAGCGCGACGCTCACAAGGTCTCTCATCGCGAGGTGAAGAGC 1348  
 Qy 430 MetThrGlyAsnAspValCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449  
 Db AGGCTGGGAATACCCAGTGTGCGACTGCGGGGCTGCAGACCCACGCTGCTCAGACCC 1408  
 Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469  
 Db AACCTGGCGGTGCTCACCTGCATCCAGTCTCGGGCGTCCACCGCGAACTGGCGTGC 1468  
 Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489  
 Db TTTTCGCGCATGCACTCACTTGCACCTTGCACCTGCGGCCCTCCGAGTGTGTGTGCGC 1528  
 Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509  
 Db TTGAACATGGGAAACAGAGCTTCAATGAGTCTATGAGGCGCCAGCTACCTTCCACCGC 1588  
 Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529  
 Db GGCCCTAAACCTCAGCTGAGAGTGACATGGGCGCCCGCAGCGAGCTACATATGGCCAAG 1648  
 Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549  
 Db TATGTGAGCATAGTGTTCACGCGG-----TGACACACCTGAGCCTCAGCGA 1696  
 Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569  
 Db CTCTGGCAGCATTTGCCAACAGGAGCCTCTCTGCTGCTGAGGCGCTTTGCCCATGGG 1756  
 Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589  
 Db CAGGACTTTGACAGCGCGCTGCCAGGCGCTGATGCACAGCGACCTGAAGAAGCTCGTCT 1816  
 Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609  
 Db CATTTGGGTGTCAAAAGTCGCCAACAGGCTTCCCTCGCTCTCTGTTGTTTTCATCATCAG 1876  
 Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629  
 Db AACGGTGTCACTGGATGCCAAGGCTCTGACGGGAAACAGGCTCTGCATCTACGACGA 1936  
 Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIle 649  
 Db CTCTACACACGACCGCGCTGCTCAAGCTGTCTGCTGAAGGGGAGAGCTTTGTTGGCACA 1996  
 Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLysLeuHisGlyCys 669  
 Db GTAAATGAAGCAGCGGAGACAGCTCTGGACATAGCAGGAAGAACACCAAGAGGTGT 2056

Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689  
 Db 2057 GAGGAGCTGCTGGAGAGCCCGAGCGGACCTTTGCTCCCTCCCTACATGAGTAC 2116  
 Qy 690 GluTTPArgLeuLeuHisGluAspLeuAspGluSerAspMetAspGluLeu 709  
 Db 2117 TCCTGGTAATTCACAGAGCTGGCTGACAGTGAGAGGATGAGGAGAGAG--- 2173  
 Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729  
 Db 2173 ----- 2173  
 Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaValGluLys 749  
 Db 2174 ----- 2203  
 Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuSer--- 768  
 Db 2204 CACTGGCCAGTGGGAGGCTGACATCAGCAACAGACCTATGAGACTGTGCCAGCCTG 2263  
 Qy 769 GlySerProProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788  
 Db 2264 GGAGCA-----GCCACCCTCAGGCGAGAGTGAGGAGCTGCCCGCCCTTGCCA 2314  
 Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuThrLysThrAsn 808  
 Db 2315 GTCAAAAC-----TCTTCTCGGACTTG----- 2338  
 Qy 809 SerValSerValAspGlySerArgGlnArgSer-----Ser-SerAspProPr 825  
 Db 2339 -----GTCAAGGGTGTCAAGACATGCGCAGTGAGATGCTTGAAGTCTCCAGC 2389  
 Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845  
 Db 2390 CTGACTTCAGAGGCCCTGAGACCCCTGAGAGCTGGGAGTCCAG----- 2435  
 Qy 845 rProProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865  
 Db 2436 -CCTCTCTCCAGTCTGATGAGCCCT-----TGGAACTGGGATCCAGCCAA 2485  
 Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProLeuProProGlnProProSe 885  
 Db 2486 GCCCCACCACCTCTGAAGAGGCTCCGAGAGCCCGGAGGACCTCCAGACCCAGCCTG 2545  
 Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905  
 Db 2546 ACAT-----CCGGGACCCCTTCGGAGATGATCC----- 2576  
 Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925  
 Db 2577 -----TCCCGCTCAGAT---TCAGCTCCGAGAGCACTCGCTCTATCGCGG 2620  
 Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerGln 940  
 Db 2621 GGGGCGGAGCCCTGAAGATGGTCCCTCAGCAGG-CAGCCTCTGCCCAAGGAAGAGCT 2679  
 Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960  
 Db 2680 GCCG----- 2683  
 Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGln 980  
 Db 2683 ----- 2683  
 Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000  
 Db 2684 -----GTTGGCATCTCAAGAGAGATGGCTCAAGAGATGGGAGTCTCCAGCAAGTTC 2736  
 Qy 1000 eValHisPheIleAlaAsp 1006  
 Db 2737 TGTGCAACTTTTGCAAGAC 2755

RESULT 15

ADK81946  
 ID ADK81946 standard; DNA; 4050 BP.  
 AC ADK81946;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Hepatocellular carcinoma up-regulated gene DDEF1L.  
 KW ds; Gene; cytostatic; antisense gene therapy; up-regulation;  
 KW hepatocellular carcinoma; antisense.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 46..2757  
 FT /\*tag= a  
 FT /product= "DDEF1L protein"  
 FT /gene= "ddef1l"  
 XX  
 PN WO2003027322-A2.  
 XX  
 PD 03-APR-2003.  
 XX  
 PF 25-SEP-2002; 2002WO-JP009873.  
 XX  
 PR 25-SEP-2001; 2001US-0324261P.  
 PR 23-AUG-2002; 2002CA-02399569.  
 XX  
 PR (UITYY ) UNIV TOKYO.  
 PA (ONCO-) ONCOTHERAPY SCI INC.  
 XX  
 PI Nakamura Y, Furukawa Y;  
 DR WPI; 2003-371927/35.  
 DR P-PSDB; ADK81947.  
 XX  
 PT New isolated nucleic acid, polypeptide encoded by it and antibody against  
 the polypeptide, for detecting hepatocellular carcinoma in a subject.  
 XX  
 PS Claim 1; SEQ ID NO 1; 89pp; English.  
 XX  
 CC The invention relates to isolated nucleic acids comprising genes that are  
 up-regulated in hepatocellular carcinoma cells; or comprising strands  
 that hybridize under high stringent conditions to the genes or a  
 complement. The nucleic acids, polypeptides encoded by them, or  
 antibodies against the polypeptide are used to detect hepatocellular  
 carcinoma in a subject. Antisense oligonucleotides to the nucleic acids  
 are used for inhibiting growth of hepatocellular carcinomas. This  
 sequence corresponds to the DDEF1L gene which encodes a protein with  
 similarity to a centaurin protein and contains an Arf GTPase-activating  
 protein domain and 2 ankyrin repeats.  
 XX  
 SQ Sequence 4050 BP; 910 A; 1163 C; 1162 G; 815 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2.13e-83 Length: 4050  
 Score: 2089.50 Matches: 455  
 Percent Similarity: 59.20% Conservative: 153  
 Best Local Similarity: 44.30% Mismatches: 275  
 Query Match: 39.63% Indels: 145  
 DB: 10 Gaps: 19

US-09-914-042-1 (1-1006) x ADK81946 (1-4050)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20  
 Db 46 ATGCGGAGCAGTTCAGCGTCCCGAGTTCCTGCGCGTCACCGCGAGGACCTCAGCTCC 105  
 Qy 21 ProThr---AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 39  
 Db 106 CCGGCTGGGCGCGCGCTTCGCCGCCAAGATGCCCGGTACCGGAGGGGCGCGCTGGCG 165

```
Qy 40 IleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLys 59
Db 166 CGGAGGAGATCTTGGAGGAGACCAAGCCATCTGTCAGAGATATAAGAGGCTGTGCGG 225
Qy 60 AlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeu 79
Db 226 GCAATCCATAGCTCCGGCCTTGGCCATGTGGAGATGAAGACGATGACCGAGAGGCCGTG 285
Qy 80 GluLysPheGlyGlyAsnCyValCyArgAspAspProAspLeuGlySerAlaPheLeu 99
Db 286 GAATCTTTAGCAACACAGCCACTGTGCCAGAACAGCCATGAGCTCCACAGGCTTCCTA 345
Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMet 119
Db 346 AACTTGGCGGTGTTCAACCGGAGGTGCTGGCTCTTCAAGAACCTGATTCAGAACTTG 405
Qy 120 AsnAsnIleLeuSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLys 139
Db 406 AACAACTGTTCTTTCCCTCGACAGTCTGATGAAGGGGACGCTGAGGGACGGTCCA 465
Qy 140 GlyAspLeuLysLysPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLys 159
Db 466 CAGGATCCAAAAACACGCTGAGAGGATGGAAGGACTATGAAGCCANAATGCCCAAG 525
Qy 160 IleGluLysGluLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179
Db 526 CTGGNAGAGGAG--CGGATCCGCCAGGTGACAGAGGGATC----- 567
Qy 180 GlyAlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCys 199
Db 568 CTTGGGAGGTGGCCAGGACATGCAGAGAGCGCGCATCTTCCAGCTGCACATGTGT 627
Qy 200 GluTyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219
Db 628 GAGTATCTGCTCAAAAGCCGGGAGGACGAGATGAAGCAAGGTCTCTGACTTCTCCAGAGC 687
Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGlu 239
Db 688 CTCATCAAGTCTTCCAGCCGACGACAACTTTTCCAGATGGTGGAGGCTGCCCAG 747
Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259
Db 748 AGCTGTGTTCCCTTCATCGAAGCTGGCGGCTCAGTACATGCATCCATCAGGCCAC 807
Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
Db 808 GAGGACGAGCTACAGAAGTACGACCTCCGGGACTCCCTCCGAGGGACACTGCAGCTT 867
Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297
Db 868 CAGACGACAGAGGAACACCTGAGCCGGAAGAACTCAGGATGTGGCTATAGCATCCACCAG 927
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
Db 928 CACCAAGGCAACACGAGCTTTGGGACGGAGAAAGTGGCTTCTATACAAGAAAGTGAC 987
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 988 GGAATTCGAAGAGTCTGGCAGAAAAAGGAAGTGTGGAGTCAAGTATGGCTGTGACCATC 1047
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 1048 TCACACACGACGATAAAACGGCCCCCGGTGAAGCTGACCTGTGCTGACGTGCCAGTGAGG 1107
Qy 358 ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
Db 1108 CCAAAACCTTGAGGAGAAAAAGTCTTCGACTGTGTGACCCCTGCTGACGTGCCAGTGAGG 1167
Qy 378 GlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 1168 CAGGACAGGAGGACGACGAGTGTGAGCGGTGGTGTGAGTGTTCGAAACAGCAAGGAC 1227
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAsp-----AspAsn 409
```

```
Db 1228 GAAGCCCTGACGACGCGCTTCTCGGGAGACCCAGCGCTGCGCGGGTCTCTGGGGTCC 1287
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 1288 GCCGCCCATGATGGGAGCCGACAGCCTCAAAAGCTGTCTCATCGCGAGGTGAAGAGC 1347
Qy 430 MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
Db 1348 AGGCTTGGGAATAGCCAGTGTCTGCGACTGCGGGGCTGCAGACCCACCGTGTCTAGCAC 1407
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 1408 AACCTGGCGGTGCTCACTGTCATCGTCTCGGGCGTCCACCGCGAACTGGCGTGC 1467
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 1468 TTTTGGCGCATGCGACTCACTCACCTTGGACCTGTGTGGGCCCTCCGAGTTGTGTGCGC 1527
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
Db 1528 TTGAACATGGGAACAACAGAGCTTCATGAGGTCTGAGAGGCCAGCTACCTCTCACCGC 1587
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 1588 GGCCCTAAACCTCAGCTGAGAGTGACATGGGCACCCCGCAGGAGTACATTTATGGCCAAG 1647
Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
Db 1648 TATGTGAGCATAGTGTTCACGCGCGG-----TGACACCTCTGAGCCTCAGCGA 1695
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
Db 1696 CTCTGGACAGCAATTTGCAACAGGGACCTCTCTGCGTACTTGGAGGCTTTGCCAATGG 1755
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
Db 1756 CAGGACTTTGGACAGCCGCTGCCAGGCGCTGATGCACAGGCACCTGAAGAACTCGTCTTG 1815
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
Db 1816 CATTTGGCTGTCAAAAGTCGCCAACAGGCTTCCCTGCTCTCTGGTGGATTTTCATCCAG 1875
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
Db 1876 AACGCTGTACCTTGGATGCCAGGCTGCTGACGGGAACACGCTCTGCATAGCAGCA 1935
Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIle 649
Db 1936 CTCTACAACACGACCGCGCTGCTCAAGCTGCTGAAGGGGAGAGCTTTGTTGGGCACA 1995
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 1996 GTAAATGAAGCAGCGAGACAGCTCTGGACATAGCCAGGAAGAAGACCAAGAGAGTGT 2055
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2056 GAGGAGCTGTCTGGAGCAGGCCCGCGGGACCTTTGCTTCCCTCTACATGTGGACTAC 2115
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2116 TCCTGGGTAAATTTCCACAGAGCCTGCTCAGAGTGAGGAGGATGAGGAAGAAG--- 2172
Qy 710 GlnProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2172 ----- 2172
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2173 -----CGCTGTGCTGAAGCTCCCGGCCGAGCT 2202
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
```

```
Db 2203 CACTGGCGCAGTGGGAGGCTGGACATCAGCAACAGACCTATGAGACTGTGCGCAGCCTG 2262
Qy 769 GlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaProLeuPro 788
Db 2263 GGAGCA-----GCCACCCCTCAGGGCGAGAGTGTCCCGCCCTTGCCA 2313
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsn 808
Db 2314 GTCAAAAAC-----TCTCTCGGACTTGT----- 2337
Qy 809 SerValSerValAspGlySerArgGlnArgSer-----Ser-SerAspProPr 825
Db 2338 -----GTCCAGGTTCGACAGATCATCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2388
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2389 CTGAGTTCCAGAGGCCCTCGACCCCTGAGAGCCTGGCGAGTCCAG----- 2434
Qy 845 rProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2435 -CCTCCTCTCCAGTCTGATGAGCCCT-----TGGAACCTGGGGATCCAGCCAA 2484
Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProLeuProGlnProProSe 885
Db 2485 GCCCCACCACTCTGAAGAGGGCCTCGAGAGCCCCCAGGACCTCCAGACCCAGCCTG 2544
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 2545 ACAT-----CCGGGACACCCCTTCGGAGATGATACC----- 2575
Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2576 -----TCCCGCTCAGAT---TCAGCTCCGAGAGCACTCGCTCTATCGCGG 2619
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 2620 GGGCGCGGAGCCCTGAAGATGGTCCCTCAGCAGG-CAGCCTCTGCCCCAGAGGAACGT 2678
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2679 GCCG----- 2682
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
Db 2682 ----- 2682
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2683 -----GTTGGCATCACTGAAGAGATGGCTCAAGGACTGGGAGTCTCCAGCAAGTTC 2735
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2736 TGTGCACTTTTGCAGAC 2754
```

Search completed: August 4, 2005, 08:58:12  
Job time : 1406 secs

**THIS PAGE BLANK (USPTO)**